ADVANCES IN BLIND DECONVOLUTION

Bruce L. K. Davey

A thesis presented for the degree of
Doctor of Philosophy
in Electrical and Electronic Engineering,
in the University of Canterbury,
Christchurch, New Zealand.

May 1989.

Abstract

Theoretical and practical aspects of image restoration and neurological signal processing are presented. All descriptions of algorithms are accompanied by examples.

All of the popular deconvolution algorithms that find application in image processing contexts are comprehensively reviewed, and inherent limitations of deconvolution are identified. Four categories of deconvolution algorithms are discussed in detail.

The first category encompasses those *conventional deconvolution* techniques that are applicable in situations where estimates of the blurring function are available.

Ensemble blind deconvolution techniques, for situations in which many differently blurred versions of a single object are available, comprise the second category of deconvolution algorithms. Astronomical speckle processing, a collection of image processing techniques in optical and infrared astronomy for high spatial resolution imaging through the Earth's atmosphere, is the field in which ensemble blind deconvolution techniques find the most widespread application. Popular astronomical speckle imaging techniques are reviewed with one such technique, shift-and-add, and extensions to it, being examined in detail. Application of shift-and-add to fields other than astronomical speckle imaging are also outlined. Zero-and-add, a new ensemble blind deconvolution technique, is described in the context of one-dimensional astronomical speckle imaging. Zero-and-add compares the complex spectral zeros of the speckle images to identify the zeros common to all of the speckle images. From the common zeros a reconstruction of the object is computed. The effect upon the zero-and-add algorithm of contamination and windowing of speckle images is investigated. Extensions to zero-and-add allowing the technique to process two-dimensional speckle images are described. A composite method for processing one-dimensional infrared speckle images, which incorporates several image processing techniques including shift-and-add and zero-and-add, is introduced and invoked to process infrared speckle images of the astrometric binary star Ross 614 AB. This method automatically compensates for deficiencies in the scanning mechanism used to record infrared speckle images, and is self-calibrating for atmospheric effects.

The problem of reconstructing an image, or the phase of its spectrum, from its known spectral magnitude is called *phase retrieval*, which characterizes the third category of algorithms. This is a special subclass of the blind deconvolution problem in which the true image and blurring function are conjugate mirror images of each other. Iterative techniques implementing both phase retrieval, and the related problem of recovering the spectral magnitude from the phase, are reviewed. The direct phase retrieval algorithm (recently reported by R.G. Lane and co-workers), based on partitioning the complex spectral zeros of a convolution into the sets of zeros of the individual components, is examined. The effect of contamination upon this algorithm is investigated.

The status of the fourth category, which are *general blind deconvolution algorithms* capable of deconvolving a single blurred image without knowledge of the

blurring function, is reviewed. A new iterative algorithm, capable of deconvolving the contaminated convolution of two components that can, in general, be complex, is introduced. The ability of general blind deconvolution algorithms to improve the image resulting from the shift-and-add algorithm is emphasised and illustrated.

The manifestation of epilepsy in the electroencephalogram of patients with suspected neurological disorders is outlined and a computerized system for the automated detection of this epileptiform activity is described. The system comprises two distinct stages. The first is a feature extractor, written in the conventional procedural language FORTRAN, which utilizes parts of previously published spike-detection algorithms to produce a list of all spike-like occurrences in the EEG. The second stage, written in the production system language OPS5, reads the list and employs rules incorporating knowledge elicited from an electroencephalographer (EEGer) to confirm or exclude each of the possible spikes. A summary of the detected epileptiform events is produced which is available to the EEGer for interpreting the EEG. The performance of the expert system is compared with an EEGer reading (in the standard neurological fashion) a 320 second segment from an EEG containing epileptiform activity.

Acknowledgements

I am indebted to many people for the advice, encouragement and support I have received during the period of my Ph.D. study, and especially during the onerous task of preparing this thesis.

In particular, I appreciate and admire the enthusiasm and direction provided to me by my supervisor Professor Richard Bates during the course of my research and the effort he expended (appreciated on some occasions more than others) attempting to expunge all gaucheries, hideosities and other abominations from this thesis.

I greatly value the support, understanding and love afforded to me by Helen Bush during the period of my Ph.D. research and I am grateful to her for the effort she made proof reading this (engineering) document. I also value the support and friendship I have recieved from flatmates, the Bush family, and members of both Saint Ninian's youth group and the 'Wrinklies' during my time as an undergraduate and postgraduate student in Christchurch.

I thank my colleagues (academic staff, technical staff and fellow students alike) at the University of Canterbury, and at Christchurch Hospital, from whom I have learnt so much through many stimulating seminars and discussions. I extend an especial thanks to Dr. Richard Fright for his insight, example, and effort.

I will always remember the many lunchtimes spent on Ilam fields for the friendly bash at soccer, and I treasure the, much too brief and infrequent, periods away from it all in the ruggedly beautiful wilds of New Zealand's South Island.

Thanks are due to: E.K. Hege, of the Steward Observatory, University of Arizona, Tucson, Arisona, USA; N. Baba, of the Applied Physics Department, Hokkaido University, Sapporo, Japan; and P. Wood, of the Mount Stromlo and Siding Springs Observatories, Australian National University, Canberra, Australia, for supplying astronomical data processed by myself, and other members of the University of Canterbury speckle group.

I acknowledge the receipt of both a New Zealand University Grants Committee Postgraduate Scholarship and the Christchurch City Council Electricity Department Research Scholarship.

Finally, I wish to sincerely thank my mother and father, and the rest of my immediate family, who, throughout my life, have loved, supported and encouraged me. Without their enduring effort and assistance this thesis could not be.

Contents

List of Figures

List of Tables

Preface

The advent of the digital computer has created an information processing revolution in most fields of scientific endeavour. As computers become increasingly powerful and rapidly reduce in cost it is becoming realistic for ever more computationally demanding problems to be routinely tackled. Attempts to obtain solutions to problems can only be made, however, when algorithms for arriving at the solution from given data exist. Furthermore, for a computer to arrive at a solution, the existing algorithm must be implementable on that computer. This thesis details development of algorithms to harness a computer's computational power to retrieve or reconstruct information which has been obscured in specific ways.

Professor R.H.T. Bates' research group at the Department of Electrical and Electronic Engineering of the University of Canterbury has, for the last 20 or so years, been active in developing algorithms for applying computers to a diverse range of topics of engineering and scientific interest (Bates 1987a). Research in fields including general inverse problems (notably computed tomography and ultrasonic imaging), astronomical image processing, satellite communications and various biomedical engineering problems have advanced as a result of efforts of members of the group under Richard(I)'s direction.

Entering this group in 1985, firstly as a master's student and later as a PhD student, afforded me a rather unusual opportunity to undertake research in two quite separate areas of information retrieval. Researching two different problems, each involving the development of computer algorithms but using very different programming techniques, allowed me to obtain a broader understanding of the possibilities and problems involved in applying computers to the task of retrieving or reconstructing information which has been obscured. This opportunity arose, for a large part, because of the presence of W.R. Fright as a postdoctoral fellow at the University of Canterbury from 1984–1987. Richard(II), who is actively interested in image processing, was then, and indeed still is, also involved with research with several departments at Christchurch Hospital, including the Department of Neurology.

The first of my research areas was in the field of neurological information processing. Neurology, as its name implies, is the branch of medicine concerned with the study of nerve systems. The brain is the organ which acts as the the body's communication and computing centre and is composed almost entirely of nerve cells, or *neurons*. (It is the interconnections of various types of neurons in the brain which provides memory and the ability to reason. Other neurons are designed to transport information between the brain and the remainder of the body.) Medical specialists (in this case *electroencephalographers* or EEGers) can glean a considerable amount of information about the function and disorders of the brain from a recording of its electrical activity. This recording is called an *electroencephalogram*, usually abbreviated to EEG. For the reader unfamiliar with electroencephalography, introductory information relevant to my research is presented in Appendix A.

It has been known for many years that epilepsy often causes characteristic elec-

trical activity on the scalp. EEGers routinely read an EEG to detect the presence of this *epileptiform* activity. The salient feature of epileptiform activity in the EEG is the intermittent *spike* waveform (Chatrian *et al.* 1974). Over the last two decades there has been considerable interest in automating this detection task and numerous 'spike detection' systems for have been reported (for reviews, see, Ktonas 1983; Gotman 1985). However, each of these systems have had only limited success (Gevins 1984; Gotman 1985). Thus, I was interested in developing an improved computer based system for the automating the spike detection process.

Recognizing the inherent suitablility of this task for solution by *expert system* techniques lead to the implementation of the detection process using a particular type of expert system known as a *production system.* The resulting system, which is described in Appendix B (*cf.* Davey *et al.* 1989a), was developed in conjunction with members of the staff of the Department of Neurology and the Department of Medical Physics and Bioengineering at Christchurch Hospital. Results of the system operating on EEGs containing epileptiform activity, as well as significant artefact, suggests that our system will usefully complement an EEGer when reading an EEG for epileptiform activity. Research directed toward implementing the system in a form suitable for routine clinical use in the Department of Neurology is continuing.

Expert systems are one of the developments resulting from research into Artificial Intelligence. They facilitate representation of the knowledge of an *expert* in a particular field within a computer in a form that the computer can use. As Artificial Intelligence methods increase in sophistication they are certain to find widespread technological and scientific application.

My second area of research, and that with which this thesis is principally concerned, involves the *deblurring* (or *deconvolution*) of images. Recorded images are frequently blurred in some way, as for example from using an out-of-focus camera, or moving a camera whilst making an exposure, or viewing an object through a randomly fluctuating propagation medium, such as the earth's atmosphere. The purpose of a deconvolution algorithm is to remove the blurring to recover a pristine version of the object that is being imaged.

Conventional deconvolution algorithms allow an image to be deblurred if a reasonably faithful estimate of the function characterizing the blurring is available. The recorded image is deblurred using a filter derived from the estimate of the blurring function. The inverse filter and the Wiener filter are examples of multiplicative conventional deconvolution algorithms (Bates and M^{c}Donnell 1986). Several subtractive deconvolution algorithms have also been proposed.

Blind deconvolution algorithms overcome the requirement, inherent in conventional deconvolution algorithms, that accurate knowledge of the blurring be available. Algorithms of this class allow an image to be deblurred without prior knowledge of either the true object or an estimate of the function causing the blurring. It is convenient to split blind deconvolution algorithms into two categories.

The first category, here called *ensemble blind deconvolution*, allows the true object to be recovered from an ensemble of differently blurred versions of that object. Astronomical speckle imaging algorithms (Bates 1982b; Dainty 1984; Roddier 1988) provide examples of ensemble blind deconvolution algorithms. These algorithms overcome the degradation induced by the earth's atmosphere when recording images of celestial objects. They have resulted from research stimulated by Labeyrie's original proposal of speckle interferometry (Labeyrie 1970).

Shift-and-add is an ensemble blind deconvolution algorithm upon which part of the research presented in this thesis is centred. SAA in its basic form (Bates 1976; Bates and Cady 1980) is both conceptually and computationally simple. However

basic SAA only produces faithful images of a limited class of objects. To extend this class, and therefore make SAA more generally applicable, numerous extensions to SAA have been proposed. Unfortunately, these extensions add significantly to the computational processing required. Thus, techniques would be welcome for extracting the true object from the SAA image for a wider class of object. Such techniques are discussed in this thesis since this is a general blind deconvolution problem. Many of the algorithms outlined in subsequent chapters are illustrated with examples of deconvolving SAA images.

ZAA (Davey *et al.* 1986) is another ensemble blind deconvolution technique for use in astronomical speckle imaging that has been developed at the University of Canterbury. Under Richard(I)'s supervision, Alastair Sinton and I jointly worked on this technique, which has also been shown to usefully complement SAA (Sinton *et al.* 1986) by providing a consistent method of deconvolving SAA images. Recently I have concentrated upon extending ZAA to allow one-dimensional speckle images taken at infrared wavelengths to be processed. The resulting composite speckle processing technique has been applied to calculating the brightness ratio and separation of the astrometric binary Ross 614 AB from infrared speckle scans (Davey *et al.* 1989b).

The second category of blind deconvolution algorithms allows the true object to be recovered from a single blurred image. Algorithms for tackling this, the *general blind deconvolution problem*, have only recently been realised. Richard(III) Lane, in conjunction with Richard(I) at the University of Canterbury, was largely responsible for the initial development of these algorithms (Lane and Bates 1987b; Lane 1988). The concept of the zero-sheet, which the three Richards jointly developed in the context of Fourier phase retrieval (Lane *et al.* 1987), provides a theoretical justification for the uniqueness of the deconvolution problem for more-than-one-dimensional images. Richard(III) also effected an algorithm based on the zero-sheet concept for the direct solution of the blind deconvolution problem (Lane and Bates 1987a). Unfortunately, this algorithm cannot, as yet, be reliably used in a practical imaging situation, where the convolution is unavoidably contaminated (Lane 1988). The theoretical justification for the uniqueness of the deconvolution process provided by zero-sheets has, however, stimulated further research into developing practical deconvolution algorithms (Bates and Davey 1988).

Previously developed deconvolution algorithms are reviewed and described in this thesis. In addition, several new algorithms of considerably increased versatility are also introduced. The contents of each of the chapters which comprise this thesis are summarised in the following paragraphs, which also identify my original research contributions.

Chapter 1 presents a brief overview of causes of degradation of measured data. It is argued that all degradation can be described by a general formula. Two special cases of this general formula which are relevant to this thesis are discussed. The first occurs when the entity that one wishes to measure is unable to be isolated from other signals impinging upon the measurement apparatus (§1.3). A practical example of an algorithm, which finds application in the field of electroencephalography (Appendix A), for overcoming this class of degradation is presented in Appendix B. The second special case of general signal degradation occurs when a single blurred signal is recorded (§1.4). It is algorithms for overcoming this class of measurement degradation with which the remaining chapters are concerned. Consequently Chapter 1 concludes with a discussion of the potential applications of these algorithms.

In Chapter 2 a mathematical basis upon which algorithms discussed in later chapters are built is presented. Notation and conventions are developed for describing the algorithms reviewed and introduced in this thesis. The Fourier and z transforms

are defined and the relationship between them discussed. The Fourier transform has proven to be an especially useful mathematical tool in image processing and is central to many of the image processing algorithms presented in this thesis. Properties of the Fourier transform which are relevant to the remainder of the thesis are collected in this chapter.

Chapter 3 comprises an introduction to deconvolution. In this chapter a few of the many algorithms that have been proposed for the conventional deconvolution (i.e. where an estimate of the blurring is available) of images or signals are reviewed. An algorithm for implementing ensemble blind deconvolution is discussed, and two inherent problems of deconvolution are introduced.

Chapter 4 provides a brief review of the most popular of the many astronomical speckle imaging algorithms which have been realised. These algorithms implement ensemble blind deconvolution, and, although they were developed originally for astronomical use, they have also found application in other fields, including ultrasonic imaging (Minard *et al.* 1985) and speech processing (Brieseman *et al.* 1987).

Algorithms for retrieving the Fourier phase (or Fourier magnitude) from knowledge of the Fourier magnitude (or Fourier phase) are reviewed in Chapter 5. These problems are known as the Fourier phase and Fourier magnitude problems. The Fourier magnitude problem fits well into the overall deconvolution theme of this thesis since it is a blind deconvolution problem. The Fourier magnitude problem is reviewed because, although not itself a deconvolution problem, it does occur in one algorithm for the solution of the blind deconvolution problem. Computational examples of reconstructions generated by applying several phase and mangitude retrieval algorithms are presented in this chapter. In addition, an original detailed pictorial study of the effect of noise on zero-sheets of an autocorrelation, and thus upon direct phase retrieval, is provided.

In Chapter 6 a new ensemble blind deconvolution technique, zero-and-add (Davey *et al.* 1986), is discussed. Zero-and-add is essentially a one-dimensional technique although its extension to two-dimensions has been achieved. The zero-and-add technique has recently been combined with several other image processing techniques to form a composite scheme for processing one-dimensional speckle images. The majority of the material discussed in Chapter 6 describes original research.

Chapter 7 is concerned with the most general, and most difficult, of deconvolution problems — deconvolving a single contaminated blurred image. The three practical algorithms that have been proposed to date (Bates and Davey 1988) are discussed. The first algorithm due to Richards (I and III) (Bates and Lane 1987b; Lane 1988) firstly applies an iterative Fourier phase retrieval algorithm and subsequently an iterative Fourier magnitude retrieval algorithm to blindly deconvolve a positive image. This algorithm has been applied in the astronomical setting (Bates and Davey 1987b). The second algorithm (Ayers and Dainty 1988) uses a single iterative loop to deconvolve a single positive image. The third algorithm (Davey *et al.* 1988a), which employs a similar loop, was under development at the same time as that of Ayers and Dainty, and constitutes original research. This algorithm is the most general of the three because it can deconvolve a single image which is, in general, complex.

Chapter 8 contains both conclusions obtained from the research into the deconvolution problem and in automated spike detection in the EEG, and my suggestions for continuing research in these fields.

My original research contributions constitute the majority of the material contained within Chapters 6 and 7 and Appendix B, although aspects of other all other chapters, with the exception of Chapter 2, are also original. The FORTRAN and

Data Type	Complex Component	Black	White
Binary mask	—	0	1
Bipolar (real)	—	minimum	maximum
Positive (real)	—	0	maximum
Complex	magnitude	0	maximum
Complex	phase	$-\pi$	π

Table 0.1: The mapping of the image data to the grey-scale for different image types.

OPS5 code comprising the software for the spike detection system presented in Appendix B was written entirely by me. The majority of the software for implementing the image processing algorithms discussed in this thesis was written by myself in either FORTRAN or PASCAL to interface to the **improc** (**im**age **proc**essing) utility (developed principally by Richard(III)). Graphical output has been produced using PLOT79 subroutines called via a software interface written by Peter Gardenier.

It is convenient here to comment upon the diagrams depicting signals (including images) found in this thesis. Three different kinds of computer generated graphics are used, namely, one-dimensional line plots, hidden-line drawings and grey-scale pictures. The grey-scale pictures have been produced using either a video printer (256 levels of grey), or a laser printer (32 levels of grey). The mapping of image data to the uniform grey-scale of the output display, for several types of image, is listed in Table 0.1. Unless otherwise specified the two-dimensional arrays are 64×64 pixels. Also, when spectra are displayed, the dc value corresponds to the image centre.

Publications and presentation prepared while the research presented in this thesis was in progress are listed below in approximate order of preparation:

DAVEY, B.L.K., SINTON, A.M. and BATES, R.H.T. (1986), 'Zero-and-add', *Optical Engineering*, Vol. 25, No. 6, June, Pp. 765–771.

SINTON, A.M., DAVEY, B.L.K. and BATES, R.H.T. (1986), 'Augmenting shift-and-add with zero-and-add', *Journal of the Optical Society of America A*, Vol. 3, No. 7, July, Pp. 1010–1017.

BATES, R.H.T. and DAVEY, B.L.K. (1987a), 'Computationally-cost-effective speckle imaging', In *Proceedings of the Joint Workshop on High-Resolution Imaging from the Ground Using Interferometric Techniques*, Oracle, Arizona, January 12–15, Pp. 219–222.

BATES, R.H.T. and DAVEY, B.L.K. (1987b), 'Towards making shift-and-add a versatile imaging technique', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol. 828: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 87–94.

DAVEY, B.L.K. and THORPE, C.W. (1987), 'Image and signal reconstruction by shift-and-add', In *IPENZ conference proceedings, Volume 2*, IPENZ, Christchurch, New Zealand, May, Pp. 147–157.

DAVEY, B.L.K., FRIGHT, W.R., CARROLL, G.J. and JONES, R.D. (1989), 'An expert system approach to detection of epileptiform activity in the EEG', *Medical and Biological Engineering and Computing*, Vol. 27. To appear.

DAVEY, B.L.K., FRIGHT, W.R., CARROLL, G.J. and JONES, R.D. (1988b), 'An expert system for detecting epileptiform activity in the EEG', *New Zealand Medical Journal*, Vol. 101, 28 September, P. 610. Abstract of presentation for the 6 April 1988 meeting of the Christchurch Medical Research Society.

DAVEY, B.L.K., LANE, R.G. and BATES, R.H.T. (1988c), 'Blind deconvolution of noisy complex-valued image', *Optics Communications*, Vol. 69, No. 5–6, 15 January, Pp. 353–356.

BATES, R.H.T. and DAVEY, B.L.K. (1988), 'Deconvolution ancient and (very) modern', In ALLOIN, D. and MARIOTTI, J.M. (Eds.), *Proceedings of the Cargèse Advanced Study Institute Conference on Diffraction Limited Imaging with Very Large Telescopes*, NATO, Kluwer Academic Publishers.

DAVEY, B.L.K., COCKE, W.J., BATES, R.H.T., M^{c}CARTHY, JR., D.W., CHRISTOU, J.C. and COBB, M.L. (1988a), 'Infrared speckle observations of binary Ross 614 AB: Combined shift-and-add and zero-and-add analysis', *The Astronomical Journal*. To appear.

Glossary of Notation

Three spaces spanned by images are invoked in this thesis. These spaces are K-dimensional and are called image-space, Fourier-space and z-space. Arbitrary points in these spaces are described by the position vectors $\mathbf{x}$, $\mathbf{u}$ and $\boldsymbol{\zeta}$ respectively. Quantities exising in these spaces are called images, spectra and z-spectra.

The following conventions are adhered to in this thesis:

(i) Vector quantities are indicated by a boldface variable, e.g. $\mathbf{x}$.

(ii) Images are identified by lower case roman symbols, e.g. $q(\mathbf{x})$.

(iii) Spectra are indicated by upper case roman symbols, e.g. $Q(u, v)$.

(iv) z-space spectra are distinguished from Fourier-space spectra in that they are functions of (ζ, γ), rather than (u, v). Thus, $Q(\zeta, \gamma)$ is the z-spectrum of $q(x, y)$.

(v) The autocorrelation of an image is identified by a repeated lower case roman symbol, e.g. $qq(x, y)$ is the autocorrelation of $q(x, y)$.

(vi) Estimated quantities are denoted by a superscipt caret, e.g. $\hat{q}(\mathbf{x})$ indicates an estimate of $q(\mathbf{x})$.

Symbols

Symbols employed in this thesis are defined below in approximate order of definition. A few of these symbols take on slightly different meanings in different chapters. This is noted in the text of the chapter, but the alternative definition is defined in glossary. Symbols that operate on an image are defined here with a general image $q(\mathbf{x})$ (or its spectrum).

Chapter 1

$\odot$	convolution

Chapter 2

$\imath$	the imaginary unit
$\lvert q(\mathbf{x})\rvert$	the magnitude of $q(\mathbf{x})$
$\mathcal{P}\left[q(\mathbf{x})\right]$	the phase of $q(\mathbf{x})$
$\mathcal{R}\left[q(\mathbf{x})\right]$	the real part of $q(\mathbf{x})$
$\mathcal{I}\left[q(\mathbf{x})\right]$	the imaginary part of $q(\mathbf{x})$
$\mathcal{E}\left[q(\mathbf{x})\right]$	the energy of $q(\mathbf{x})$
ϵ_{dB}	the signal to noise ratio of an image

c_r	the amount of correlation of the contamination
$L_{q(\mathbf{x})}$	the extent of $q(\mathbf{x})$
$B[q(\mathbf{x})]$	the image-box of $q(\mathbf{x})$
$S[q(\mathbf{x})]$	the support of $q(\mathbf{x})$
$\mathbf{\Delta}$	the sampling interval
$\delta(\mathbf{x})$	the delta function, or impulse
$\int(K)\int$	K-dimensional integration
$\star$	correlation
$\mathcal{A}[q(\mathbf{x})]$	the autocorrelation of $q(\mathbf{x})$
$\mathcal{F}[q(\mathbf{x})]$	the Fourier transform of $q(\mathbf{x})$
$\mathcal{F}^{-1}[Q(\mathbf{u})]$	the inverse Fourier transform of $Q(\mathbf{u})$
$\longleftrightarrow$	Fourier transform pair
$p^{\phi}_{q(\mathbf{x})}$	the projection at angle ϕ of $q(\mathbf{x})$
$S^{\phi}_{Q(\mathbf{u})}$	the slice of Fourier-space at angle ϕ of $Q(\mathbf{u})$
$\mathcal{Z}[q(\mathbf{x})]$	the z-transform of $q(\mathbf{x})$
$\mathcal{Z}^{-1}[Q(\zeta)]$	the inverse z-transform of $q(\zeta)$

Chapter 3

$\Omega(\mathbf{u})$	inverse filter
$W(\mathbf{u})$	Wiener filter
$\ln[Q(\mathbf{u})]$	the cepstrum of $Q(\mathbf{u})$
$\langle\cdot\rangle_M$	the ensemble average of
$\Xi[Q(\mathbf{u})]$	the contamination mask of $Q(\mathbf{u})$
$E[q(\mathbf{x})]$	the expected value of $q(\mathbf{x})$

Chapter 4

D	telescope diameter
θ	angular resolution of telescope
r_0	fried seeing parameter
τ_s	short exposure duration
λ	wavelength
$\Delta\lambda$	wavelength bandwidth
λ_0	centre wavelength in band
R	brightness ratio of binary star
ρ	separation of binary star
θ	position angle of binary star
$\mathbf{s}$	vector separation of a binary star
$\boldsymbol{\alpha}$	vector displacement for KT processing
$\odot^{-1}$	deconvolution

Chapter 5

Υ	the set of points satisfying constraints
E_T	true error
E_I	image-space error
E_F	Fourier-space error
$Z\{Q(\zeta)\}$	the set of zeros of $Q(\zeta)$, or, equivalently, the zero-sheet of $Q(\zeta)$

Chapter 6

$w = u + \imath v$	generalization of the variable, u, over the complex plane
$Z\{Q(w)\}$	the set of zeros of $Q(w)$, or, the zero-map of $Q(w)$, or, a pixellated image representing the zero positions of $Q(w)$
$\bar{Z}\{Q(w)\}$	the averaged zero-map of $Q(w)$
$\bar{Z}_T\{Q(w)\}$	the thresholded averaged zero-map of $Q(w)$

Chapter 7

$E_{I,i}(q)$	i^{th} image-space error of $q(\mathbf{x})$
$E_{T,i}(q)$	i^{th} true-space error of $q(\mathbf{x})$
$q_{\Xi}(\mathbf{x})$	$q(\mathbf{x})$ bandlimited by $\Xi[Q(\mathbf{u})]$
γ	support overestimation parameter
α	filter constant of modified Wiener filter

Abbreviations

Selected abbreviations that appear in this thesis are:

CAA	correlation shift-and-add
dB	decibels
DFT	discrete Fourier transform
EEG	electroencephalogram
EEGer	electroencephalographer
EFET	entire function of exponential type
EMG	electromyogram
FFT	fast Fourier transform
GS	Gerchberg-Saxton
KT	Knox-Thompson
LHS	left hand side
LWH	Lynds-Worden-Harvey
psf	point-spread-function
psv	point spread variant
RHS	right hand side
SAA	Shift-and-add
SNR	Signal to noise ratio
SSW	spike or sharp wave
TC	Triple correlation
ZAA	Zero-and-add

Glossary of Astronomical Terminology

A glossary of astronomical terms that are used in the chapters relating to astronomy is presented here. This material is drawn from Hedley Robinson (1972), Barlow (1975) and Roy and Clarke (1977)

Asteroid - A small body orbiting the Sun, usually between the orbits of Mars and Jupiter.

Binary star - A system of two stars which are located sufficiently near in space to be connected by the bond of mutual graitational attraction, compelling them to orbital about their common centre of mass.

Double star - A system comprising two stars. Some double stars are optical doubles merely because its components happen to lie closely in line of sight without having any real physical connection (*cf.* binary star).

Infrared radiation - Electromagnetic radiation at wavelengths longer than that of red light but shorter than radio waves.

Limb - The region near the edge of the visible disc of an object.

Limb darkening - A falling off of brightness near the limb of an object.

Magnitude system - The brightness of a star, planet etc., measured according to a logarithmic scale in which a difference of five magnitudes defines a brightness ratio of 100 to 1.

Multiple star - A star made up of more than two components.

Nebula - A mass of tenuous gas and dust in space.

Planets - Non-luminous bodies in orbit around a luminous primary body.

Position angle - The apparent direction of one object from another, measured from the primary through East from North.

Primary mirror - The principle mirror of a reflecting telescope.

Primary star - The star, of a multiple star, that is chosen as the reference star. It is usually the brightest of the component stars.

Relative magnitude - The ratio of two stellar brightnesses, B_1 and B_2 are related to their magnitudes, m_1 and m_2, by

$$\frac{B_1}{B_2} = 2.512^{-(m_1-m_2)}.$$

Scintillation - The rapid variations of apparent brightness (i.e. twinkling) occurring when observing an object through the earth's atmosphere.

Secondary mirror - The smaller mirror which faces the primary mirror of a reflecting telescope, and which focuses the final image at a convenient distance behind the primary mirror.

Secondary star - The component of a double star that is not chosen as the primary star. It is usually the less bright of the two.

Seeing - The degrading effect on image quality which results from the transient optical properties of the atmosphere in the line of sight.

Spectroscopic binary - A binary star having components too close to be observed visually, but shown as a binary by period variation in Doppler shifts as the stars circuit around each other.

Supergiant star - A very large luminous star.

Visual binary - A binary star that can be resolved as comprising two components either by eye at the telescope, or by subsequent processing of recorded images.

Chapter 1

Retrieval of Obscured Information

Scientific investigation is critically dependent upon one's ability to make accurate measurements of whatever *entities* are of interest. Each entity depends upon the area of investigation and can be as diverse as, for example, the distance of a star from Earth, the wavelength of an electromagnetic wave, or the intensity of sound emitted by a jet aircraft on take-off. Measurements of the entity are made and the resulting data are subsequently analysed. Conclusions drawn from such analysis are important for obtaining an improved understanding of our world. Advance in science can be considered as a process of proposing a theory, performing experiments involving the measurement of some entity or entities of interest, and from the information resulting from processing these measurements evaluating the range of applicability of the proposed theory. If the theory survives examination then it remains until such time as measurements of higher quality are made or old measurements are processed in new ways allowing new conclusions to be reached. These conclusions may reveal discrepancies in the theory, leading to either a modification of the theory, or its abandonment.

All measurements are made using a recording instrument of some kind. Instruments may simply consist of the human senses or may incorporate sophisticated (and usually expensive) devices dedicated to measuring the particular entity of interest. Any measuring device introduces an uncertainty or error when data are recorded, the extent of the error largely being determined by the quality of the instrument. All analyses of measured data should attempt to extract the required information from the data but the conclusions drawn must recognize the limitations of the measuring device.

A further complication often arises from it being impossible, or impractical, to make direct measurements of the entity of interest. This may occur because the entity is inaccessible for some reason, allowing only remote recordings to be made. A striking example of this is provided by astronomy. It is not (as yet) feasible to send instruments to investigate objects outside our solar system. So, in Science's ongoing quest to understand the universe, measurements can only be made at a great distance from the object. Another situation of great practical importance involving inaccessible entities is the probing of internal organs of the human body. Ingenious schemes for non-invasive imaging of body tissue are now available for use by clinicians (Kak 1979; Wells 1982). These schemes continue to be the subject of considerable research (Bates *et al.* 1983; King and Moran 1984).

The inability to make direct measurements may, alternatively, occur when it is only possible to make measurements of an entity appreciably different from, but of course related in some definite way to, the entity of interest. Often, by appropriately processing the recordings, information about the entity of interest can be

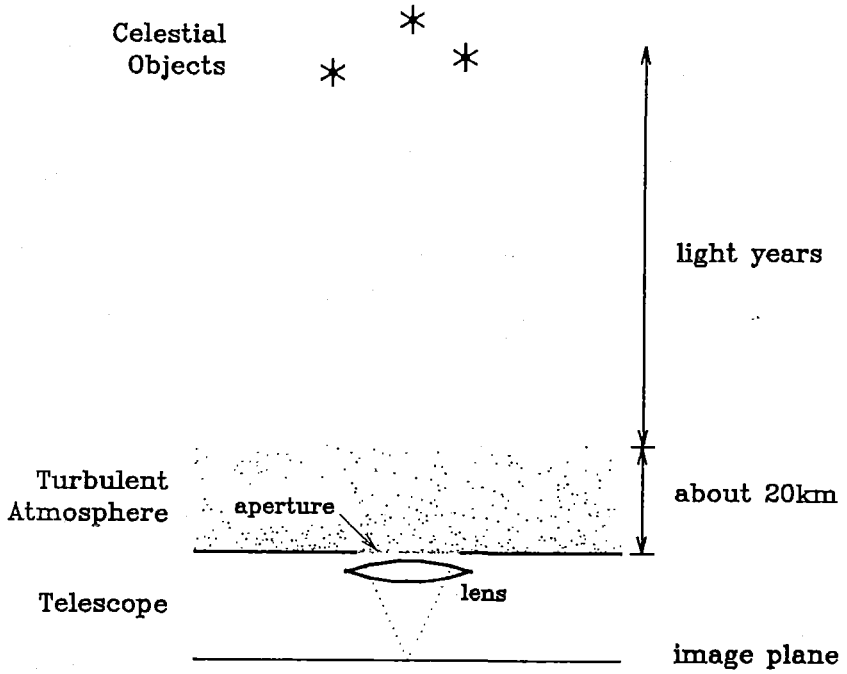


Figure 1.1: Schematic diagram depicting optical astronomical imaging with an earth-based telescope. The telescope records radiation created by, or reflected from, distant celestial objects. This radiation must pass through the relatively thin, but turbulent, atmosphere of the earth which produces a severe degradation of the recorded signal.

reconstructed. For instance, the highest resolution observations in radio astronomy are made with synthesis telescopes, which are a type of interferometer (Thompson *et al.* 1986). Consequently, the measurements yield Fourier-space information concerning celestial objects that radiate at radio wavelengths. Inverse Fourier transforming data obtained from such measurements allows the object being imaged to be reconstructed (Cole 1977; Christiansen and Högbom 1969).

The imprecision of practical astronomical instruments, and the unavoidable remoteness of the object being measured, both cause an inevitable degradation of the recorded information. Many similar causes exist in other fields of science. The purpose of this chapter is to provide a unifying description of such degradation before providing a rationale for pursuing the development of algorithms for deblurring signals and images, the central theme of subsequent chapters.

Different fields of signal processing have developed specific terminology to describe degrading processes. Terminology employed in this thesis is introduced and defined in §1.1. A general approach that encompasses all degraded signals is introduced in §1.2. Two special cases of the general signal, for which techniques for signal recovery are established, and which have particular relevance to this thesis, are identified and discussed in §1.3 and §1.4. The chapter concludes with a section briefly introducing techniques for overcoming the blurring of signals. The various classes of deblurring algorithms are introduced in §1.5 and potential practical applications for such processing are mentioned.

The entities discussed in this thesis are, in practice, recorded as functions of either time or space. A function of time is usually referred to as a *signal* and necessarily has a single dimension. A function of space is usually called an *image* since a particularly common form of this type of information is the two-dimensional picture or image.

Images can be of one, two, or more dimensions, although those discussed in this thesis are usually limited to one or two dimensions. For ease of exposition, both signals and images are referred to by the more general term signal in this chapter. However, the term image appears more often in subsequent chapters that are predominantly concerned with image analysis. The dimensionality of the function representing the signal of interest, if not explicitly stated, is always apparent from the context in which the function appears.

Notation, abbreviations, and conventions adhered to when discussing algorithms in this thesis are listed in the Glossary which starts on page xxi. A further glossary that begins on page xxv lists terminology invoked when describing astronomy and astronomical image processing. It is appropriate to include this since many of the examples illustrating the algorithms introduced in this thesis relate to the field of astronomy.

1.1 Describing Signal Degradation

In many different fields of technological endeavour one has occasion to measure signals that have been degraded. Terminology to describe both the signal degradation and the signal recording processes have been established in all these fields. This section introduces and briefly describes terms that appear in the following sections and chapters of this thesis. Since this thesis is principally concerned with rather general applications of image processing, emphasis in placed on terminology employed in this field. However, where appropriate, terminology relevant for describing the specialized medical applications introduced in Appendices A and B is included.

Object: the particular entity about which information is required.

Recording apparatus: the equipment that measures, and records, information about the object.

Optical system: recording apparatus designed to gather electromagnetic radiation within the wavelength region extending from ultraviolet (40nm) to the far infrared (1mm).

Emanation: physical process (e.g. radiation) emitted by, or scattered from, the object of interest.

Detector: the part of the recording apparatus where the actual recording is made. In an optical system the detector is typically photographic film, or a light sensitive electronic device.

Propagation medium: the space through which the emanations propagate before reaching recording apparatus.

Aberration: a defect in the recording apparatus, causing the recorded signal to be imperfect. Common aberrations in an optical system include spherical and chromatic distortion and astigmatism. Born and Wolf (1970, §§ 5 and 9) present a detailed discussion of aberrations in such systems.

Noise: spurious recorded information that is not part of the signal. Noise can be introduced by many sources at various stages of the signal measurement process. For instance, the statistical nature of photon arrival when imaging at very low light levels and thermal noise introduced by the recording instrument are both examples of noise commonly occurring in many optical systems.

Contamination: a general term encompassing noise and any other departures from an ideal recording situation. The term also includes departures of the recording system, and the propagation medium, from the model upon which the system is based. For example, departures from the isoplanatic model (see §1.4) are incorporated.

Band-limited recording: a measurement situation in which only a finite range, or band, of frequencies are recorded. Usually it is the lower frequencies that are faithfully recorded.

Ideal recording apparatus: recording apparatus that is aberrationless i.e. it is perfectly aligned. It is, however, here taken to be band-limited.

Resolving power: a measure of the ability of the recording instrument to form separate and distinct signals of two objects close together. Because of diffraction, no optical system can form a perfect reconstruction of a point source, but produces instead a small disk of light (the Airy disk) that is surrounded by alternately dark and bright concentric rings (called Airy rings).

Diffraction limited imaging: an imaging situation employing ideal recording apparatus. Furthermore, the propagation medium introduces no distortion.

Unresolvable object: an object of sufficiently small angular extent that it has no features resolvable by the recording instrument. Such an object can be considered to be a point source and is recorded as an Airy disk.

True signal: the signal that would be measured by ideal recording apparatus in the absence of any degradation processes. The true signal is, however, here understood to be band-limited.

Blurring: a specific type of degradation of a measured signal. Blurring occurs in situations in which the recording situation is imperfect in some sense. For example, if an optical system has some figuring error (e.g. spherical aberration), or is out-of-focus, the recorded signal is blurred. Other situations involving blurring arise due to a limitation imposed by the propagation medium through which the instrument views the objects of interest. Examples occur when recording signals through a turbulent medium, such as the atmosphere of the earth, or the ocean.

Point-spread-function (psf): the function characterizing the blurring of a signal.

Artefact: a rather general term describing imperfections arising in the signal recording and analysis process. The term is often employed in medical signal processing and medical imaging contexts. The causes of such imperfections may be separated into three major categories:

- Components of the recorded signal that are not part of the true signal. Such artefacts may arise from many sources, for example, limitations in the recording system or extraneous signals being detected by the recording instrument.
- Errors occurring when assumptions upon which the instrument is based are no longer valid. A typical example of such artefacts occurs in ultrasonic imaging where the assumptions of constant propagation speed and uniform refractive index throughout the medium are often significantly in error and lead to false information in the recorded signal.

- Apparent, but incorrect, structure introduced by the processing performed on the measured data. An example of this is provided by the basic shift-and-add algorithm (see §4.8) when it processes data from certain classes of object. Here false structures (called ghosts) are induced in the reconstructed signal by the processing.

Reconstructed signal: a version of the true signal that results after applying restorative processing to the measured signal.

1.2 The General Degraded Signal

A general, contaminated and blurred, recorded signal $s(\mathbf{t})$ can be expressed as the summation of individual signals, each of which are in general differently blurred, i.e.

$$s(\mathbf{t}) = \sum_j \mathcal{L}_j \{f_j(\mathbf{t}), h_j(\mathbf{t})\} + c(\mathbf{t}), \tag{1.1}$$

where $\mathbf{t}$ is a vector uniquely identifying a position in *space-time* and $\mathcal{L}_j$ is some arbitrary linear operator acting upon the j^{th} signal $f_j(\mathbf{t})$ and the j^{th} blurring function $h_j(\mathbf{t})$. $c(\mathbf{t})$ is a function incorporating any contamination in the recording process that cannot be modelled by $\mathcal{L}_j$. All practical signal recording situations introduce contamination. Usually, the corruption caused by the contamination can not be separated from the true data, so that the conclusions derived from the measurements are constrained in accuracy by the level of the contamination.

Signal processing problems require a recorded signal to be processed in some way to obtain a reconstructed signal. The processing usually involves isolating a particular feature of the recorded signal. Thus, invoking the notation defined in (1.1), it is appropriate to pose the **General Signal Processing Problem:**

"Recover $f_j(\mathbf{t})$ for a particular value of j, given $s(\mathbf{t})$."

The generality implicit in (1.1) means that it is impossible to construct an algorithm for solution, in all instances, of the general signal processing problem. Practical solutions can only be obtained when the general problem is customized for the particular measurement situation.

Consider a hypothetical situation, depicted in Fig. 1.2, in which it is required to isolate the sound produced by a single whale belonging to a large school of similarly sounding whales. Furthermore, the apparatus that is recording the sound is located a considerable distance from the whale of interest. In this situation the sound emitted by each whale is degraded because of multiple reflections off the whales, the sea's surface and any nearby obstacles. Thus, a blurred version of the sound of each whale impinges upon the detector. Consequently the task of extracting a 'deblurred' version of the 'call' of a particular whale is extremely difficult.

On a somewhat more intricate scale, consider the difficulty in isolating the component of the electrical activity, induced on the external surface of the body, by a single one of the many nerve cells controlling the heart. Here, the electrical signal emitted from each neuron is degraded by the passage through the tissue between the heart and skin. Thus, a superposition of degraded signals impinges on the detector. Some of these signals originate in the heart, but many are also due to other sources of electrical activity in the body, for example, the neurons controlling muscle tissue. It is, in effect, impossible to isolate the signal produced by a single neuron in such a degraded recording.

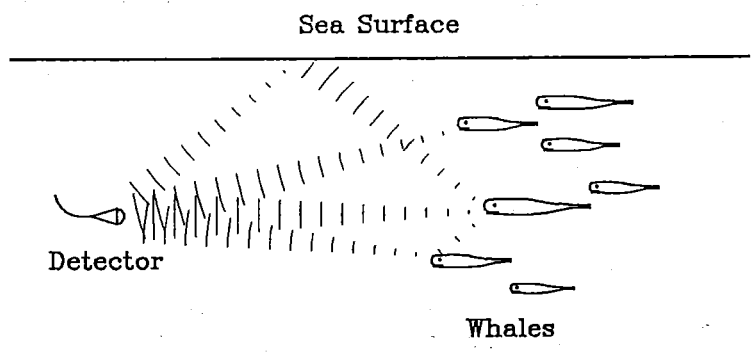


Figure 1.2: The task of isolating the sound by a particular whale in a large school of whales is extremely difficult since many blurred versions of the whales' sounds impinge upon the detector.

By making the problem more specific, through dealing with special cases, the problem of recovering a particular $f_j(\mathbf{t})$ often becomes tractable. Many important practical signal recording situations exist in which the signal degradation can, at least to a good approximation, be expressed as such a special case. Two such special cases, which are of particular interest to this thesis, are introduced in the following sections.

1.3 Intermingled Signals

Consider a situation in which several signals impinge upon the recording apparatus, but it is desired to isolate only one of them. Often the sources of the signals are located in such a way that it is impossible to position the detector to prevent the extraneous signals from being detected. Furthermore, in the general signal recording situation described by (1.1), each of the signals impinging on the detector is blurred in some unknown manner.

Consider now the special case of (1.1) in which none of the individual signals are blurred. It follows that $\mathcal{L}_j$ and $h_j(\mathbf{t})$ in (1.1) disappear and the resulting composite signal, $s(\mathbf{t})$, can be expressed as a summation of the *intermingled* signals $f_j(\mathbf{t})$, i.e.

$$s(\mathbf{t}) = \sum_j f_j(\mathbf{t}) + n(\mathbf{t}). \tag{1.2}$$

Thus, in this situation, the general signal processing problem involves separating the required signal from the extraneous ones.

The task of separating the intermingled signals that comprise the recorded signal can usually only be accomplished when specific characteristics of the individual constituent signals are known. For example, if it is known that the Fourier transforms of the component signals occupy different frequency bands, it is possible to separate the components by appropriately filtering the signal (Oppenheim and Schafer 1975, Chapter 10). Another situation arises when *a priori* knowledge about the shape of characteristic features of the waveform are known. Pattern recognition techniques can then be applied, allowing the individual signals to be identified and subsequently separated (Batchelor 1978).

Examples of typical situations in which the signal of interest is obscured by additional, extraneous signals are listed in Table 1.1. Examination of this table reveals that this type of obscuration of the true signal typically occurs when measurements

Recording situation	Entity of interest	Typical extraneous signals
Electrocardiography	Electrical activity of the heart (ECG)	All other electrical signals on the skin
Electroencephalography	Electrical activity of the brain (EEG)	Electrical signals on the scalp not originating in the brain
Radio astronomy	Stellar radio sources	Background cosmic radiation
Radar	Signal reflected from a particular object	Reflections from additional targets (clutter)

Table 1.1: Typical signal recording situations in which signal analysis is complicated by extraneous signals impinging on the measuring instrument.

of an intricate system are being made. The problem is aggravated when, in addition, measurements must be made at a considerable distance from the source of interest, i.e. they are made remotely.

Such a situation arises, for instance, in radio astronomy. Here an earth-based radio telescope is employed to capture electromagnetic radiation at radio wavelengths emitted from stellar objects. However, in addition to the radiation originating in the object of interest (i.e. the true signal), the universe has background radiation that also impinges upon the telescope. Thus, in the recording, an extraneous signal is intermingled with the signal of interest.

The human body is another example of an intricate system containing many inaccessible 'components'. Here, the components are typically internally located organs. Since many of these organs are critical for the healthy functioning of the body, there is considerable interest in the measurement of the signals emanating from them for diagnostic purposes. To make direct measurements of internal organs, invasive techniques, which usually require surgery, are required. Unfortunately, this places a patient at considerable discomfort and often at risk. Thus, techniques allowing the information to be extracted from externally recorded, but necessarily intermingled, signals would be welcome.

An example of an algorithm for retrieving and processing an intermingled signal originating in the body is described in Appendix B. The medical field in this measurement situation is neurology and the recording is of the electrical activity the brain. This signal is usually recorded by means of electrodes placed on the scalp and is called an electroencephalogram (EEG) (see Appendix A). The electrical voltages induced on these electrodes by the activity of the brain are recorded on a paper chart, which is subsequently 'read' by the EEGer. Since the duration of a typical recording is 20–30 minutes, the EEGer must 'process' a very large amount of information when reading an EEG. Thus, a considerable amount of time and effort is expended by the EEGer. Some means of reliably automating the EEG analysis task would be desirable, both to relieve clinician fatigue and increase uniformity of the conclusions reached in the reading process (Ktonas 1983; Gotman 1985).

One particular signal that is often required to be isolated in the EEG is well known to be characteristic of a patient suffering from epilepsy (Spehlmann 1981; Gevins 1984). Such *epileptiform* activity is intermingled with the other ongoing neurological signals that, although containing considerable potentially useful information, have little or no relevance to the specific detection task. The particular algorithm presented in Appendix B was developed by the author to automate the process

Situation	Cause of blurring	Blurring Type
Speech processing (Markel and Gray, Jr. 1976)	Filtering caused by vocal tract	psi
Photography	Out-of-focus camera	psi
	Relative movement between the object and the camera	psi
Astronomy (Barlow 1975) Ultrasonic imaging (Abbott and Thurstone 1979; Bates and Minard 1984)	Object viewed through a turbulent medium	psv/psi
Seismology (Ulrych 1971)	Non-ideal impulse propagated into earth's crust	psi

Table 1.2: Typical situations of technological importance requiring the reconstruction of a single blurred recorded signal. In several of these fields the exact nature of the recording situation determines whether the blurring can be considered psi or psv.

of identifying this epileptiform activity. The algorithm operates in two distinct stages. The first stage is a feature extractor that outputs a list of features describing activity that has the shape of that expected of epileptiform activity. The second stage inputs the list of features and applies a rule based expert system to ascertain which of the features are truly epileptiform. Preliminary results have demonstrated the system's potential for detecting and classifying epileptiform activity.

1.4 Blurred Signals

A further special case of the general degraded signal arises when a single blurred signal $b(\mathbf{t})$ impinges upon the measuring instrument. Since there is only a single signal (i.e. $j = 1$), it is appropriate to remove the subscript j and the summation from the general equation. Thus, such a signal can be expressed as

$$b(\mathbf{t}) = \mathcal{L}\{f(\mathbf{t}), h(\mathbf{t})\} + c(\mathbf{t}). \tag{1.3}$$

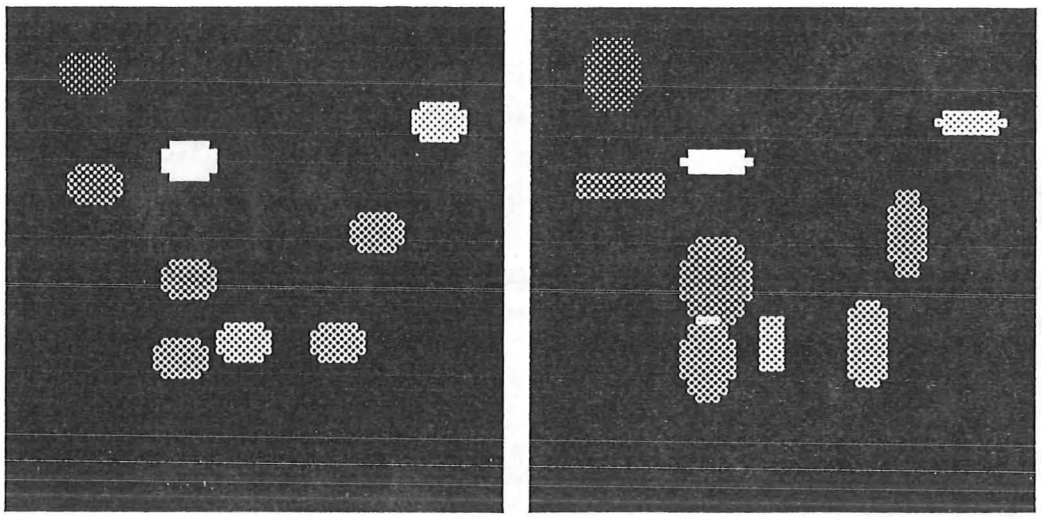
The general linear operator $\mathcal{L}$ and the psf $h(\mathbf{t})$ together completely characterize the blurring. There are a multitude of situations of scientific interest in which a single blurred signal is recorded. Representative examples of such situations are listed in Table 1.2.

Isoplanatic or *point-spread-invariant* (psi) blurring occurs when the psf is the same for the entire signal. When this is true, the linear operator in (1.3) reduces to *convolution* (denoted here by $\odot$). Thus, the isoplanatically blurred signal is described by

$$b(\mathbf{t}) = f(\mathbf{t}) \odot h(\mathbf{t}) + c(\mathbf{t}). \tag{1.4}$$

Perfect isoplanatic blurring seldom (if ever) occurs in practice. However, in many situations the deviation from isoplanatism is limited, so the system can be considered as being effectively isoplanatic. Departures of the actual system from the isoplanatic model are readily incorporated into the contamination term, $c(\mathbf{t})$, which is added to the ideally blurred signal $f(\mathbf{t}) \odot h(\mathbf{t})$ in (1.4).

When the point-spread-function (psf) causing the blurring varies appreciably throughout the recorded signal, the isoplanatic assumption is no longer valid. Such



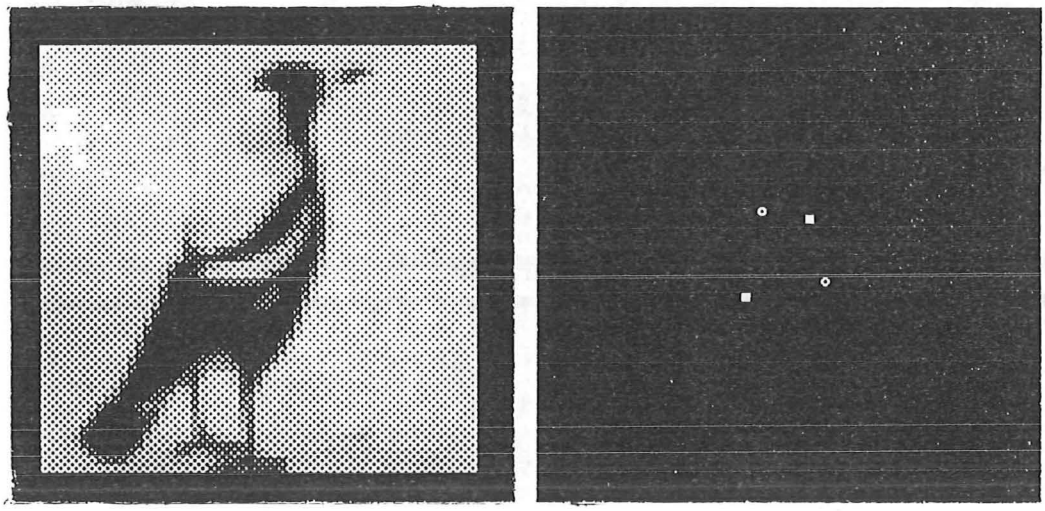
(a) (b)

Figure 1.3: Isoplanatic and nonisoplanatic blurring: (a) an isoplanatically blurred signal; (b) the same true signal blurred nonisoplanatically.

blurring is called *point-spread-variant* (psv) or *nonisoplanatic*. Fig. 1.3 displays computer generated examples of both an isoplanatically blurred and a nonisoplanatically blurred signal. The true signal has been chosen simple enough that one can immediately see the degradation caused by the blurring. Note that, in the isoplanatically blurred signal Fig. 1.3(a), each of the isolated points of the true signal has been blurred identically. However, in Fig. 1.3(b) each of these points has been blurred differently, indicating that this is nonisoplanatic blurring.

For the special case of a single blurred signal, the general signal processing problem posed in §1.2 becomes one of removing the blurring. For nonisoplanatic blurring, considerably more information must be available to overcome the blurring than for isoplanatic blurring. Thus, practical algorithms for overcoming nonisoplanatic blurring are, as yet, rarely realizable. Therefore, discussion in this thesis is restricted to those situations where the blurring can be considered isoplanatic. Consequently, the term blurring is hereafter understood to mean isoplanatic blurring unless it is explicitly stated otherwise.

Blurring occurs in many situations of considerable importance. For example, in speech processing applications it is often appropriate to consider a speech waveform to be a blurred version of an excitation function which is produced by the vocal chords (Markel and Gray, Jr. 1976). The blurring can be usefully assumed isoplanatic throughout short intervals having durations of a few (e.g. no more than 3) pitch periods. So, each short segment of a recorded speech signal is the convolution of the excitation function and the *vocal-tract response*. This convolution relationship has important practical implications for many speech processing tasks, including speech recognition and speech synthesis, since it often allows the excitation function and the vocal tract response to be separated. Another example of blurring is provided by seismic exploration in situations where an explosion creates a pulse of seismic energy that propagates through the earth. By measuring the reflections of the wave off various layers of the earth's crust, information about the earth's composition can be obtained (Ulrych 1971). However, this required information is blurred by a function describing the nonideal shape of the seismic wave. Image processing is



(a) (b)

Figure 1.4: True signals for image processing examples presented in this chapter: (a) the true signal for Fig. 1.5; (b) the true signal for Fig. 1.6.

also much concerned with situations involving isoplanatic blurring. The majority of the subsequent discussion of blurring in this thesis is directed toward describing and overcoming the blurring of images arising in various technical and scientific applications.

Four particularly common forms of functions causing blurring in image processing applications, have been identified by Bates and M^{c}Donnell (1986). These psfs, and the blurring each produces, are now discussed with reference to Figs. 1.4, 1.5 and 1.6.

When a photograph is taken with an out-of-focus camera, the resulting signal can be considered to be the true signal convolved with a psf that characterizes the out-of-focusness of the lens. For many types of camera lenses, this psf can be approximated by a uniform disk whose diameter is related to the severity of the out-of-focusness. Such a psf is depicted in Fig. 1.5(a). The blurring caused by this psf when the true signal shown in Fig. 1.4(a) is viewed, is apparent in Fig. 1.5(b). Note that the psf shown in Fig. 1.5(a) is symmetric, as all psfs describing out-of-focusness are expected to be. This has important practical implications for algorithms designed to remove out-of-focus blurring, as described in Chapter 7.

Another type of blurring is evident when relative movement occurs between a camera and an object while making an exposure. This psf would arise, for instance, when a photographer 'bumps' the camera while making an exposure, or when the camera is held stationary and the object, say a rapidly flying bird, moves. In such situations the recorded signal is blurred by a psf characterizing the motion. For a constant linear motion the psf is a straight line of uniform amplitude. An example of such a psf and the blurring it causes are provided by Fig. 1.5(c)) and Fig. 1.5(d).

The third form of psf is a collection of random impulses, which commonly arises when emanations from the object must pass through a turbulent propagation medium to reach the detector. Provided the time taken to make the recording is sufficiently short that the turbulence is effectively frozen, the recorded signal can often be usefully approximated as the true signal convolved with a psf consisting of random impulses. This psf can be considered to characterize the medium for that recording. Such a psf, and a recorded signal that has been degraded by it, are depicted in Fig. 1.6(a)

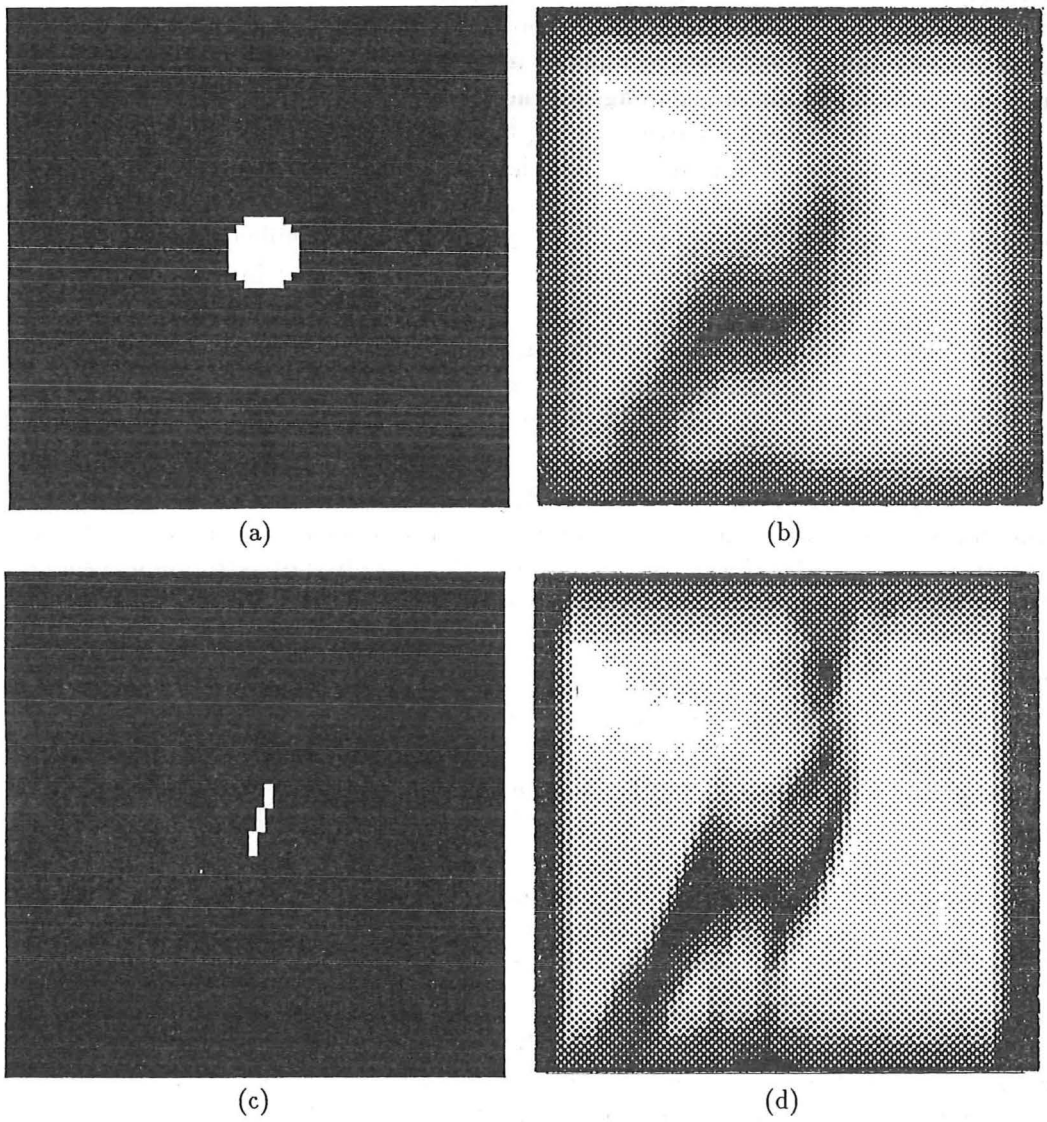


Figure 1.5: Examples of isoplanatic blurring typically occurring in photography. The true signal being recorded in these examples is shown in Fig. 1.4(a): (a) when an exposure is made with an out-of-focus camera the psf describing the lens is a uniform disk; (b) each point in the object is blurred by the psf to produce the degraded signal; (c) the psf characterizing a particular motion of the camera (or object) whilst an exposure is being made; (d) the blurred signal resulting from the psf, (c).

and Fig. 1.6(b).

There are many practical measurement situations in which signals are required to be recorded through a fluctuating medium. Consider, for example, the randomly fluctuating atmosphere of the earth. When observing celestial objects with a earth-based telescope (see Fig. 1.1) the radiation necessarily passes through this medium. Consequently, the recorded signal is severely blurred and details of the true signal are not immediately apparent in the recorded signal. This astronomical imaging situation, and algorithms for overcoming its inherent blurring, are discussed further in Chapters 4 and 6. Astronomical imaging is not the only signal measurement situation in which fluctuations in the earth's atmosphere degrade a recording. Consider, for example, the image recording situation in which, say, a zoologist is attempting to take a long-distance photograph of a desert animal. When the atmosphere of the earth is cool and still, very little degradation of the recorded signal is evident. However, when the earth's surface is heated by the sun, movement is induced in the lower regions of the atmosphere. This phenomenon is commonly known as heat haze and causes a blurring of the recorded signal. Fluctuating media also degrade signal measurement processes in fields other than image processing. For example, the random impulse psf also arises when performing acoustic imaging in a fluid media, especially water (see §4.8.4).

The fourth form of psf identified by Bates and McDonnell (1986) is similar in form to a gaussian function. Again, many examples of such a psf arise in practice. For instance, reconsider the situation involving the recording of astronomical signals with an earth-based telescope. Suppose now, however, that instead of making short duration exposures, the telescope is operated conventionally i.e. it tracks the object across the sky averaging the incident signal for many seconds, minutes, or even hours. The resulting long exposure signal can be considered to be the true signal convolved with a gaussian type psf. An example of such a psf and the blurring it causes are shown in Fig. 1.6(c) and Fig. 1.6(d). Comparing the blurred signal Fig. 1.6(d) with the true signal Fig. 1.4(b), reveals that the blurring has destroyed most of the detail apparent in the true signal.

In some situations it is possible to record a sequence, or *ensemble*, of differently blurred versions of the true signal. Consider a signal recording situation in which the blurring function (i.e. $h(\mathbf{t})$) changes with one of the dimensions of the space-time vector, $\mathbf{t}$, while the true signal $f(\mathbf{t})$ is invariant for changes of $\mathbf{t}$ in that dimension. Such a situation, that of recording short exposure images through a turbulent medium, has already been discussed in this section. In this situation the medium changes with time, while the object being observed is temporally invariant. It follows that it is thus possible to record an *ensemble* of differently blurred versions of $f(\mathbf{t})$ by making a series of measurements at different times. Invoking the integer m to identify members of the ensemble, a typical recorded signal can be specified as $b(\mathbf{t}, m)$. If the ensemble contains, say, M members then

$$b(\mathbf{t}, m) = f(\mathbf{t}) \odot h(\mathbf{t}, m) + c(\mathbf{t}, m)\,; \quad m = 1, 2, \ldots, M, \tag{1.5}$$

where $h(\mathbf{t}, m)$ and $c(\mathbf{t}, m)$ are the psf and the contamination for the m^{th} member of the ensemble.

Due to significant advances made in recent years, it is now often possible to recover the true signal from an ensemble of isoplanatically blurred versions of that signal without knowledge of the psfs that cause the blurring in each of the individual signals. Even more recent advances have enabled the true signal to be recovered from a single blurred version. These advances allow deconvolution to be much more

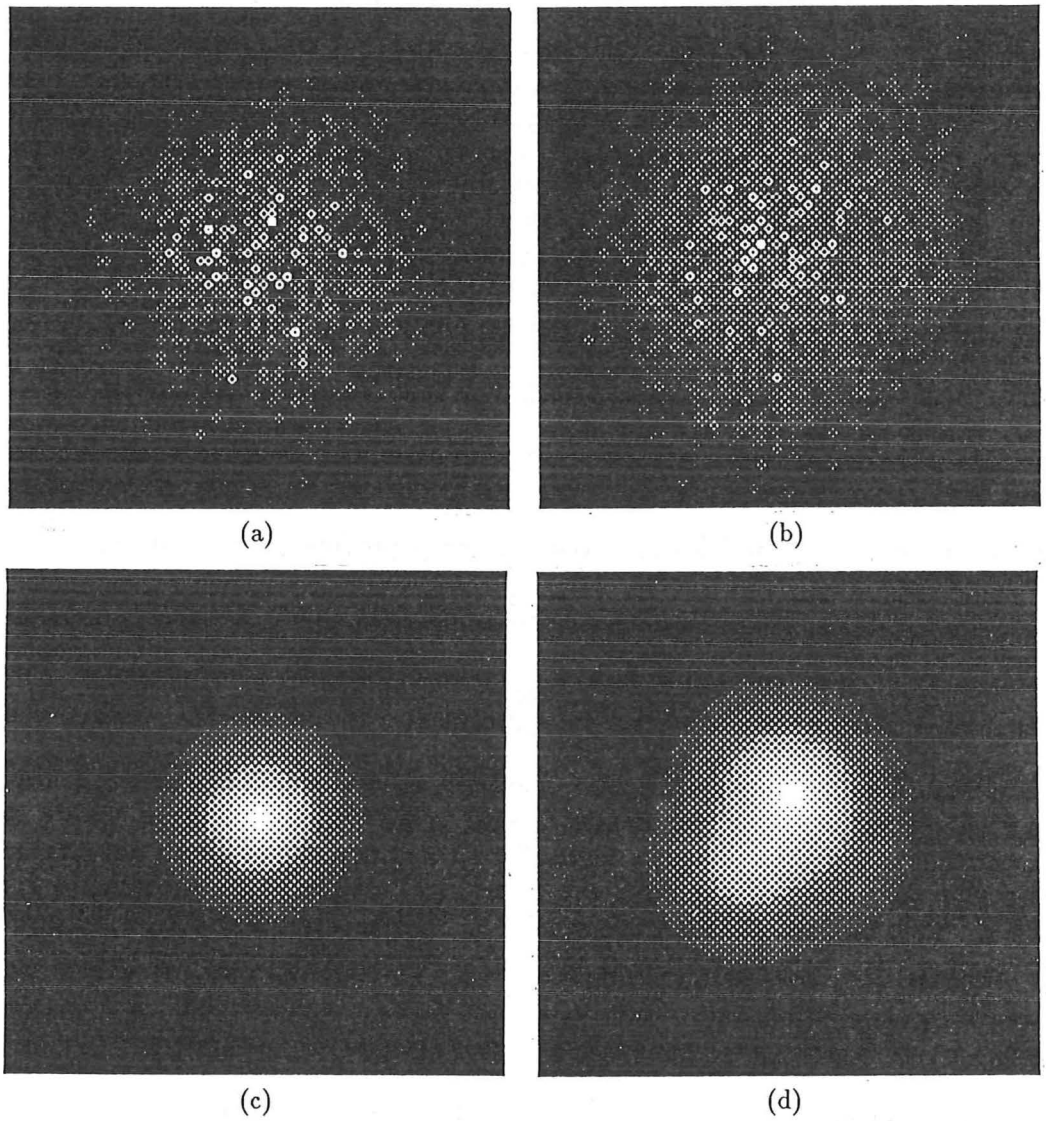


Figure 1.6: Examples of isoplanatic blurring commonly occurring when viewing an object through a turbulent propagation medium. The true signal for these examples is shown in Fig. 1.4(b): (a) a psf describing the blurring evident in an exposure made with sufficiently short duration that the turbulence is effectively frozen. Such a psf consists of a collection of random impulses; (b) the resulting blurred signal; (c) a gaussian psf typifying the blurring apparent when a long exposure image is made through a turbulent propagation medium; (d) the resulting blurred signal that exhibits virtually none of the detail apparent in the true signal.

generally applicable than it has traditionally been, since a blurred signal can be deblurred with less, and often no, *a priori* information about the nature of the blurring. In §1.5 several deconvolution problems are posed, and the various classes of algorithms that have been devised for implementing deconvolution are briefly introduced. Section 1.5 also indicates in which subsequent chapters of this thesis these algorithms are discussed in detail.

1.5 Reconstruction of Blurred Signals

This section introduces and defines terminology for categorizing techniques for overcoming the blurring (i.e. the deblurring) of signals. Since the isoplanatic blurring assumption is taken to be valid the blurring can be expressed as a contaminated convolution. Consequently, the process of deblurring a signal is often termed *deconvolution.* Classes of deconvolution algorithms are here discussed using the notation of (1.4) and (1.5).

The *conventional deconvolution* problem is posed as:

> "Given a blurred signal $s(\mathbf{t})$, and an estimate of the psf $h(\mathbf{t})$, recover $f(\mathbf{t})$."

Here, since the psf $h(\mathbf{t})$ which causes the blurring is provided, a filter allowing the blurring to be removed can be derived from it. Algorithms to implement conventional deconvolution are reviewed in Chapter 3. Some of the many proven applications of these algorithms include speech processing (Markel and Gray, Jr. 1976), communications engineering (Haykin 1983) and image processing (Andrews and Hunt 1977).

The *blind deconvolution* problem is posed as:

> "Given only $s(\mathbf{t})$ (or an ensemble of such signals), recover $f(\mathbf{t})$."

Since the form of the blurring psf is not available, conventional deconvolution algorithms can not be applied. Because an estimate of the psf is not needed for algorithms that implement blind deconvolution, they have much wider potential application than algorithms for conventional deconvolution. It is useful to further subdivide the blind deconvolution problem into two categories.

Ensemble blind deconvolution algorithms operate on an ensemble of differently blurred versions of the true signal and are thus posed as:

> "Given an ensemble of differently blurred versions of the true signal $s(\mathbf{t}, m)$; $m = 1, \ldots, M$, recover the true signal $f(\mathbf{t})$."

Such an ensemble of differently blurred signals is described by (1.5). By appropriately processing the ensemble it is often possible to recover $f(\mathbf{t})$. Ensemble blind deconvolution algorithms are discussed in §3.2 and Chapter 4 and a new algorithm and extensions to it are discussed in Chapter 6. These algorithms find application, for instance, in astronomical imaging (Bates 1982b; Roddier 1988), ultrasonic imaging (Bates and Minard 1984) and speech processing (Brieseman *et al.* 1987).

In the *general blind deconvolution* problem on the other hand, only a single blurred signal is available. Thus, it is posed as:

> "Given only $s(\mathbf{t})$, recover $f(\mathbf{t})$."

General blind deconvolution, as it is posed here, is the most general class of deconvolution problem considered in this thesis. The conventional deconvolution problem and the ensemble blind deconvolution problem can both be considered as specializations

of this problem. General blind deconvolution is also the most difficult class of deconvolution problem to solve, since one is presented with the least amount of information. Algorithms capable of achieving general blind deconvolution have only recently been reported (Bates and Davey 1988) (*cf.* Lane 1988; Ayers and Dainty 1988; Davey *et al.* 1988a) and are typically iterative in nature. They have been derived mainly from the various methods for solving the Fourier phase problem that are discussed in Chapter 5. A discussion of general blind deconvolution algorithms is presented in Chapter 7. Potential practical applications for these algorithms arise in many fields of image processing (Bates and McDonnell 1986). Typical examples are astronomical imaging (Bates and Davey 1988), photography (Bates and McDonnell 1986), crystallography (Ramachandran and Srinivasan 1970) and electron microscopy (Saxton 1978; Hawkes 1980).

Chapter 2

Preliminaries

In this chapter are grouped various concepts, terminology and mathematical tools that are fundamental to the presentation of the algorithms found in subsequent chapters. This grouping is intended to provide a mathematical and conceptual framework upon which the algorithms described throughout this thesis can be developed. Grouping these concepts in a single chapter provides the reader with a convenient reference source when reading subsequent chapters.

Concepts that are applied in general image processing applications are introduced in §2.1. Complex numbers, vectors, sampling, pixellated images, image energy, image support and image-form are discussed. The concept of mathematical operators that transform data from one 'space' to another is also introduced. Many of the variables and symbols that appear consistently throughout this thesis are also defined in this section.

The linear operations of convolution and correlation are discussed in §2.2. The autocorrelation operation, or the correlation of a function with its complex conjugate, is also introduced in this section.

The Fourier transform (Bracewell 1978; Bates and M^{c}Donnell 1986), a uniquely invertable operation that produces the spectrum of a signal, is introduced in §2.3. The theory behind this transformation was originally introduced (in primitive form) by, and was subsequently named after, Joseph Fourier (1768–1830) following his investigation of how periodic functions could be represented in terms of sine and cosine functions (Kreysig 1979). The Fourier transform has proven to be extremely useful in many areas of science, one being image recovery, as this thesis testifies.

Properties of the Fourier transform that are utilised in this thesis are discussed in §2.4. For example, the convolution theorem reveals that the convolution of two images can be obtained by taking the Fourier transforms of the images, multiplying these Fourier transforms together and inverse Fourier transforming the result. Hence, the computationally expensive operation of direct convolution is reduced to the operations of Fourier transformation and multiplication, that are usually much less computationally demanding.

When computing the Fourier transform of a sampled signal, a special form of the transform, called the discrete Fourier transform (DFT) (Stanley 1975, Chapter 9), is conveniently invoked. Considerable advances have been made in algorithms for implementing the DFT on a digital computer with the most popular of these algorithms being the fast Fourier transform (FFT) algorithm (Cooley and Tukey 1965; Ramirez 1985). The DFT and FFT algorithms, and practical implications implicit with their usage are discussed in §2.5.

The z-transform is another invertable relation that can usefully be employed to assist with the analysis of signals that are sampled at equispaced intervals (Bracewell

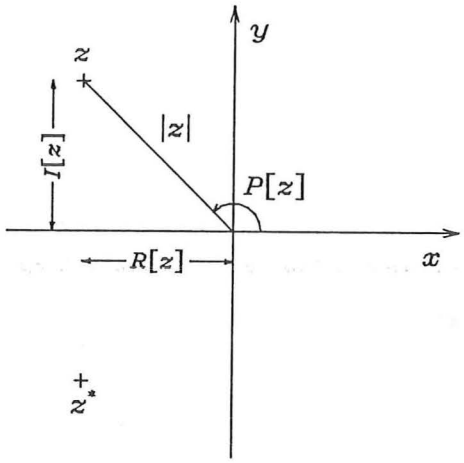


Figure 2.1: The complex plane showing the general complex variable z and its complex conjugate z^*.

1978, pp. 257 ff.). This transform and its relationship to the Fourier transform are discussed in §2.6.

Entire functions are a particular class of functions with properties that can usefully be exploited in many applications (Requicha 1980) since it transpires that the Fourier transforms of practical signals can be considered to be entire functions. A useful property of entire functions is that they are completely characterized by their (complex) zeros. This property is applied extensively in this thesis, so §2.8 is devoted to introducing these functions.

Much of the material presented in this chapter can be found in many books that cover details of Fourier theory (*cf.* Goodman 1968; Bracewell 1978; Kreysig 1979; Bates and M^{c}Donnell 1986).

2.1 Imaging Concepts

Complex numbers have been considered a useful tool by mathematicians and scientists alike, for several centuries since the concept was first introduced by Girolamo Cardano (1501–1576) to solve algebraic equations that had non-real roots (Kreysig 1979).

The concept of the complex plane is useful to aid in visualising complex numbers (see Fig. 2.1). This plane is analogous to the one-dimensional number line of real numbers. A position on the one-dimensional number line is specified by a single number. However, to specify a point on the two-dimensional complex plane, a pair of real numbers is needed. Consider the general complex variable z. The location of z on the complex plane can be uniquely specified by a pair of real variables (x, y). Invoking the imaginary unit $\imath$, which is defined to be one of the solutions to $\sqrt{(-1)}$, z can be written $z = x + \imath y$. The Cartesian coordinate values x and y are termed the real and imaginary parts of z, here denoted by $\mathcal{R}[z]$ and $\mathcal{I}[z]$ respectively, and are both real numbers within the range $-\infty$ to ∞.

Alternatively, the complex number $z = x + \imath y$ can be expressed as a cylindrical polar coordinate, $z = re^{(\imath\theta)}$, where $r = x^2 + y^2$ is a positive real number and $\theta = \tan^{-1}(y/x)$ is a real number with a principal value in the range $-\pi$ to π. r and θ are termed the magnitude and phase of z, and are here signified by $|z|$ and $\mathcal{P}[z]$

Space	Dimension (K)		
	1	2	> 2
Image	$q(x)$	$q(x,y)$	$q(\mathbf{x})$
Fourier	$Q(u)$	$Q(u,v)$	$Q(\mathbf{u})$
z	$Q(\zeta)$	$Q(\zeta,\gamma)$	$Q(\zeta)$

Table 2.1: Notation to describe an arbitrary function in image-, Fourier- and z-space.

respectively. Thus, z can be defined either by

$$z = x + \imath y = \mathcal{R}[z] + \imath\mathcal{I}[z] \tag{2.1}$$

or by

$$z = re^{\imath\theta} = |z|e^{\imath\mathcal{P}[z]}. \tag{2.2}$$

The complex conjugate of z is defined as $\mathcal{R}[z] - \imath\mathcal{I}[z]$ and is here denoted by an superscript asterisk appended to the variable. Thus,

$$z^* = \mathcal{R}[z] - \imath\mathcal{I}[z] = x - \imath y = |z|e^{\imath\mathcal{P}[z]} = re^{-\imath\theta}. \tag{2.3}$$

An arbitrary variable z, and its complex conjugate z^*, are plotted on the complex plane depicted in Fig. 2.1.

Vector quantities are denoted throughout this thesis by bold face variables. For example, the position vector identifying an arbitrary point in image-space is written $\mathbf{x}$. The dimensionality of the vector, or equivalently of space, is apparent from the context in which it appears. When dimensionality is explicitly specified, it is denoted by the integer K. When $K = 1$ (i.e. one-dimensional space) the position vector $\mathbf{x}$ can be replaced by the single variable x. When $K = 2$, the pair of independent variables (x, y) may be specified (see Table 2.1).

When processing scientific data it is often useful to invoke mathematical transformations. Such processing is said to transform the data from one space to another. Two such transformations, which find extensive application in this thesis, and which are described in subsequent sections in this chapter, are the Fourier transform and the z-transform. The spaces spanned by the data after application of these transformations are here termed *Fourier-space* — sometimes alternatively called *visibility-space* in interferometric contexts (Fright 1984) or *reciprocal-space* by crystallographers (Ramachandran and Srinivasan 1970), and *z-space* respectively. The position vectors in Fourier-space and z-space are here denoted $\mathbf{u}$ and $\boldsymbol{\zeta}$. Again, when $K = 1$ or 2 it is often convenient to replace the vector with its component variables. The corresponding notation is listed in Table 2.1.

The true object (§1.1) is signified by the function $f(\mathbf{x})$ throughout this thesis. An estimate of an image is denoted by the same variable as the image but is adorned with a circumflex accent or hat. Thus, an estimate of the true object $f(\mathbf{x})$ is written $\hat{f}(\mathbf{x})$.

2.1.1 Sampling

To be represented in a digital computer, a continuous signal must be sampled (Bracewell 1978, Chapter 10). The process of sampling is thus inherent in all digital processing. Since the algorithms discussed in this thesis are all implemented with the

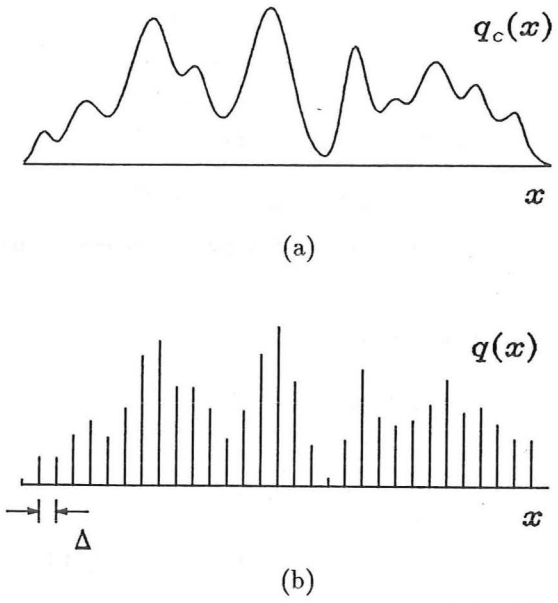


Figure 2.2: Sampling of signals: (a) A continuous bandlimited one-dimensional function $q_c(x)$; (b) $q(x)$, the signal resulting from sampling $q_c(x)$ with a sampling function having a sampling interval Δ.

aid of a digital computer, all signals are assumed to be sampled unless it is explicitly stated otherwise.

Usually, a series of equispaced samples of the signal is recorded, or specified, to give an approximation to the signal (see Fig. 2.2). The spacing between samples is here called the *sampling interval* and is denoted by the vector $\boldsymbol{\Delta}$. The components of $\boldsymbol{\Delta}$, i.e. $\Delta_1, \Delta_2, \ldots, \Delta_K$, specify the sampling interval in each Cartesian direction. The reciprocals of these components, i.e. $1/\Delta_1, 1/\Delta_2, \ldots, 1/\Delta_K$, are called the *sampling frequencies* (or *sampling rates*) in the components' Cartesian directions. Provided the sampling frequency in each coordinate direction is above the *Nyquist frequency* (i.e. twice the maximum frequency present in the signal), the samples contain all the signal information (Stanley 1975, §3.3). If the Nyquist sampling criterion is not met then signal information is lost due to aliasing (Bracewell 1978, Chapter 6).

The signals that are discussed most often in this thesis are functions of space and are usually called images. It is therefore appropriate to introduce terminology commonly employed when discussing images and image processing in the following subsections.

2.1.2 Pixellated Images

The term *pixel* (a contraction of 'picture element') is invoked to describe the element that is the basic volume unit of an image. Pixel is most straightforwardly understood in two-dimensional image processing applications where it is a two-dimensional area. However, its use is extended in this thesis to also include the basic volume unit of an image of any dimensionality. No confusion is caused by also using the term pixel to denote the value of the image within the volume element.

In general, each pixel of an image is a complex number. Such an image is said to be complex-valued, or simply complex. A real-valued image, or real image, is

a special case of the complex image in which the imaginary part of every pixel is identically zero. Thus, the phase of a real image is restricted to the values of 0 or π. A non-negative real image (called a positive image) is a special case of a real image in which the real part of every pixel is restricted to having a non-negative value. Consequently the phase of a positive image is everywhere zero.

Most naturally occurring radiating objects are *spatially incoherent* in the sense that the emanations arising at separate locations within the object do not exhibit significant *correlation* (Born and Wolf 1970, §7.1). Therefore, each of these emanations can be considered to be an independent signal. When imaging such objects it only makes sense to record the time average of the intensity impinging on the imaging instrument's detector. Thus, the recorded images are positive (Bates and McDonnell 1986, §3).

Algorithms designed to process positive images find widespread application since recordings of spatially incoherent objects are often made. However, plenty of practical situations exist where objects are both *spatially* and *temporally coherent.* In these situations phase information is needed to fully characterize signals. This information can often be inferred directly, for example, in ultrasonic imaging (Wells 1982; Minard *et al.* 1985) and radio astronomy (Thompson *et al.* 1986). Thus, algorithms capable of processing complex images also have many potential practical applications. Since a positive image is a special case of a complex image, algorithms for processing complex images can be regarded as considerably more general than those for processing positive images.

2.1.3 Image Energy and Signal-to-Noise Ratio

The energy of the image $q(\mathbf{x})$ is defined by

$$\mathcal{E}\left[q(\mathbf{x})\right] = \int\!(K)\!\int |q(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x}, \tag{2.4}$$

where $\int\!(K)\!\int$ signifies K-dimensional integration and the term $\mathrm{d}\mathbf{x}$ implies the K-dimensional volume element. When no integration limits are specified it is understood that the integration is to occur over the entire K-dimensional space.

To realistically model a practical imaging situation with the aid of a digital computer, it is essential to account for contamination which can often be modelled as being additive. Thus, letting a superscript tilde denote the contaminated version of an image, $\tilde{q}(\mathbf{x})$ is written

$$\tilde{q}(\mathbf{x}) = q(\mathbf{x}) + c(\mathbf{x}), \tag{2.5}$$

where $c(\mathbf{x})$ is the additive contamination. The level of contamination is specified in decibels (dB) by

$$\epsilon_{\mathrm{dB}} = 10 \log_{10} \frac{\mathcal{E}\left[c(\mathbf{x})\right]}{\mathcal{E}\left[q(\mathbf{x})\right]}. \tag{2.6}$$

Pseudo-random noise, which is uniformly distributed between two constant values a and b ($a \leq b$), is used to model contamination in examples of image processing presented throughout this thesis. Contamination is added to the three image-classes introduced in §2.1.2 in the following manner:

(i) For situations in which $q(\mathbf{x})$ is real, the value b is chosen and, unless otherwise specified, a is set to be $-b$.

(ii) When $q(\mathbf{x})$ is positive, it is unrealistic for the contaminated image to be negative anywhere. Thus, the same procedure as that adopted for real images is applied, except that all negative pixels of $\tilde{q}(\mathbf{x})$ are set to zero.

(iii) When $q(\mathbf{x})$ is complex, $c(\mathbf{x})$ is also complex, having both a random magnitude (uniformly distributed in the range 0 to b, where b is a positive constant) and a random phase (uniformly distributed in the range $-\pi$ to π).

It is often realistic to take the contamination to be independent from pixel to pixel. In some applications, however, the contamination recorded in adjacent pixels can not be considered to be independent, i.e. it is correlated. This *correlated contamination* is modelled in several of the simulations presented in subsequent chapters of this thesis. The modelling is achieved by, first, generating uncorrelated contamination (as described in the previous paragraph), Fourier transforming it, windowing (i.e. low-pass filtering) the result and, finally, inverse transforming. The 'amount' of correlation in every coordinate direction is here specified by the parameter c_r, which is defined by

$$c_r = \frac{u_{\max,k}}{u_{\text{cutoff},k}}, \quad k = 1, \ldots, K \tag{2.7}$$

where $u_{\max,k}$ is the maximum spatial frequency present in the k^{th} coordinate direction of the uncorrelated image and $u_{\text{cutoff},k}$ is the cutoff frequency of the low-pass filter in that coordinate direction. When specifying correlated contamination the two parameters, ϵ_{dB} and c_r, are quoted.

2.1.4 Image Support

The *support* of $q(\mathbf{x})$, here denoted by $S[q(\mathbf{x})]$, is the region outside of which

$$|q(\mathbf{x})| < \epsilon, \tag{2.8}$$

and inside of which

$$|q(\mathbf{x})| \geq \epsilon, \tag{2.9}$$

where ϵ is some positive constant small enough to be deemed negligible. When this region is finite the image is said to have finite support. In practice a recorded image is contaminated and ϵ is most appropriately chosen to be the root-mean-square (rms) magnitude of the contamination.

The *image-box* $B[q(\mathbf{x})]$ of the image $q(\mathbf{x})$ is the rectangular box, with sides parallel to the chosen Cartesian coordinate axes, that just encloses $S[q(\mathbf{x})]$. The image-box is always larger than, or equal to, $S[q(\mathbf{x})]$ (Bates 1982b). Thus,

$$S[q(\mathbf{x})] \subset B[q(\mathbf{x})], \tag{2.10}$$

where $\subset$ denotes 'is a subregion of'. The *extent* of the image in the k^{th} coordinate direction is defined as the length of the image-box in that direction. It is here denoted by $L_q(x_k)$. Fig. 2.3 illustrates the concepts of support, image-box and extent of an arbitrary two-dimensional image.

All real-world images effectively have a finite extent in each coordinate direction. They also have finite amplitude (Requicha 1980), where finite amplitude means that the image satisfies

$$|q(\mathbf{x})| < \infty \quad \text{for all} \quad \mathbf{x}. \tag{2.11}$$

Such an image is called *compact* (Fright 1984, p. 13).

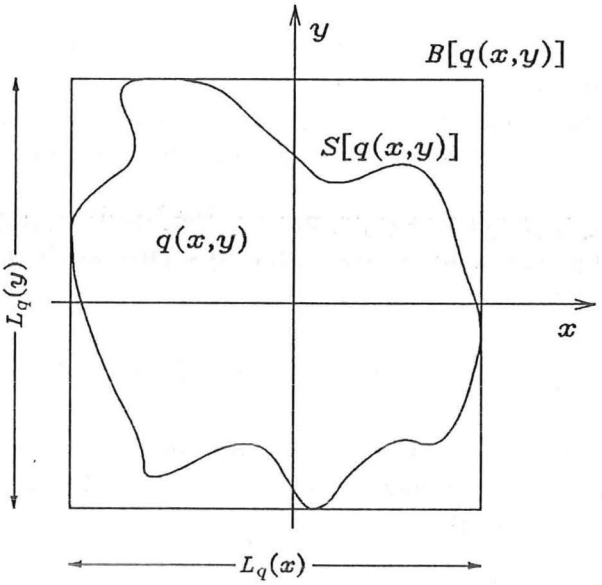


Figure 2.3: The support, image-box and extent of an arbitrary two-dimensional image $q(x, y)$.

2.1.5 Image-Form

When an image is shifted from one location in image-space to another, the image remains the same, i.e. its 'form' is unchanged. Similarly, the appearance of an image is not altered by making it brighter or fainter, or by reflecting it in the origin of image-space (i.e. rotating the image half a turn) and reversing its phase. Consider the general complex image $q(\mathbf{x})$ and the arbitrary constants $k_1, k_2, \mathbf{x}_1$ and $\mathbf{x}_2$. The scaled and translated image $k_1 q(\mathbf{x} - \mathbf{x}_1)$ and the scaled, translated, reflected and conjugated image $k_2 q^*(-\mathbf{x} - \mathbf{x}_2)$ are said to have the same *image-form* as $q(\mathbf{x})$ (Bates and M^{c}Donnell 1986, §20).

The concept of image-form is useful because, in many situations of importance in image processing, it can be reconstructed when other information about the image can not be recovered. For instance, the absolute brightness of an image may not be recoverable in situations where the image-form, which contains the relative brightness of the image's pixels, can be readily reconstructed.

2.2 Linear Systems

A *linear system* has the property that the response from several simultaneously applied inputs (stimuli) produce the same result as the summation of the responses from each of the stimuli applied individually (Goodman 1968, p. 4). This, the *linearity principle*, can be expressed as (Goodman 1968, p. 18)

$$\mathcal{L}\left[k_1 f_1(\mathbf{x}) + k_2 f_2(\mathbf{x})\right] = k_1 \mathcal{L}\left[f_1(\mathbf{x})\right] + k_2 \mathcal{L}\left[f_2(\mathbf{x})\right], \tag{2.12}$$

where $\mathcal{L}[\cdot]$ is the linear system transformation that maps the stimulus $f(\mathbf{x})$ to the output $s(\mathbf{x})$, i.e.

$$s(\mathbf{x}) = \mathcal{L}\left[f(\mathbf{x})\right]. \tag{2.13}$$

If the system is linear then (2.12) is true for all inputs $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ and for all constants k_1 and k_2. Alternatively, a linear system can be viewed as a system in

which parameters characterizing the system are not dependent upon the nature or level of the stimuli (Stanley 1975, p. 2).

Linear systems are important for two reasons. Firstly, many physical phenomena are found experimentally to exhibit the linearity principle (Goodman 1968). Secondly, mathematical tools for handling such systems are well developed. One such tool, which finds widespread application in signal processing contexts, is the *delta function.*

The *Dirac delta*, or *impulse*, $\delta(\mathbf{x})$ is a function effectively having a non-zero value (which is in fact infinite) for only a single value of $\mathbf{x}$ (Bracewell 1978, §5). It can be defined by the integrals

$$\int_{T+} \delta(\mathbf{x})\, d\mathbf{x} = 1 \quad \text{and} \quad \int_{T-} \delta(\mathbf{x})\, d\mathbf{x} = 0, \tag{2.14}$$

where T^+ is an infinitisimally small region containing $\mathbf{x} = 0$ and T^- is all regions not containing T^+. One of the many useful properties of the function is the *sifting property* (Goodman 1968, p. 18)

$$\int\!\!\!\int_{(K)} f(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}')\, d\mathbf{x} = f(\mathbf{x}') \tag{2.15}$$

which shows that any image can be considered as a sum of weighted delta functions e.g.

$$f(\mathbf{x}) = \int\!\!\!\int_{(K)} f(\mathbf{x}')\delta(\mathbf{x} - \mathbf{x}')\, d\mathbf{x}'. \tag{2.16}$$

The response of the system to a single translated delta function is called the *point-spread-function* (psf) and is defined as

$$h(\mathbf{x}, \mathbf{x}') = \mathcal{L}\left[\delta(\mathbf{x} - \mathbf{x}')\right]. \tag{2.17}$$

If the system is linear the output can be obtained by integrating the response to the individual delta functions. Thus, if $g(\mathbf{x})$ is the output of the linear system,

$$\begin{aligned} g(\mathbf{x}) &= \mathcal{L}\left[f(\mathbf{x})\right] \\ &= \mathcal{L}\left[\int\!\!\!\int_{(K)} f(\mathbf{x}')\delta(\mathbf{x} - \mathbf{x}')\, d\mathbf{x}'\right] \\ &= \int\!\!\!\int_{(K)} f(\mathbf{x}')\mathcal{L}\left[\delta(\mathbf{x} - \mathbf{x}')\right]\, d\mathbf{x}' \\ &= \int\!\!\!\int_{(K)} f(\mathbf{x}')h(\mathbf{x}, \mathbf{x}')\, d\mathbf{x}'. \end{aligned} \tag{2.18}$$

An important subset of linear systems are those for which $\mathbf{x}$ and $\mathbf{x}'$ are of the same dimension and the psf depends only upon the difference between $\mathbf{x}$ and $\mathbf{x}'$ (Goodman 1968, p. 19), i.e.

$$h(\mathbf{x}, \mathbf{x}') = h(\mathbf{x} - \mathbf{x}'). \tag{2.19}$$

Such systems are called point-spread invariant or isoplanatic (see §1.4). Substituting (2.19) into (2.18) yields the convolution integral (Bracewell 1978, p. 243)

$$g(\mathbf{x}) = \int\!\!\!\int_{(K)} f(\mathbf{x}')h(\mathbf{x} - \mathbf{x}')\, d\mathbf{x}', \tag{2.20}$$

or, when invoking the explicit convolution notation,

$$g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x}). \tag{2.21}$$

In practice no imaging system is exactly isoplanatic over the entire plane. For comparatively small departures from exact isoplanatism it is appropriate to assume that the system is effectively isoplanatic, with the departures lumped into a contamination term (which also includes measurement noise — see §1.1), i.e.

$$g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x}) + c(\mathbf{x}). \tag{2.22}$$

However, when the departures from isoplanatism are large it is often useful to split the plane into smaller regions, called *isoplanatic patches* (Goodman 1968, §2.2), within which the assumption of isoplanatism is usefully valid.

The functions f(x) and h(x) in (2.21) are said to be the *components* of the convolution. A function that cannot be expressed as the convolution of two or more components is said to be *irreducible* (Bates and Lane 1987b). It is here understood that $\delta(\mathbf{x})$ is not classed as a component function, since a function is not altered by convolution with $\delta(\mathbf{x})$ i.e.

$$g(\mathbf{x}) = g(\mathbf{x}) \odot \delta(\mathbf{x}). \tag{2.23}$$

A function that is not irreducible is said to be *composite* (Lane and Bates 1987a). A function, $g(\mathbf{x})$, that is the convolution of N components, $f_n(\mathbf{x})$; $n = 1, ..., N$, i.e.

$$g(\mathbf{x}) = f_1(\mathbf{x}) \odot f_2(\mathbf{x}) \odot \ldots \odot f_N(\mathbf{x}) \tag{2.24}$$

is here said to be N-composite.

The correlation integral is similar to the convolution integral and is defined by (Bates and McDonnell 1986, §7)

$$\begin{aligned} a(\mathbf{x}) &= \int\!(K)\!\int f(\mathbf{x}')h(\mathbf{x}+\mathbf{x}')\,\mathrm{d}\mathbf{x}' \\ &= f(\mathbf{x}) \star h(\mathbf{x}), \end{aligned} \tag{2.25}$$

where $\star$ denotes K-dimensional correlation and $a(\mathbf{x})$ is the resulting correlated function. A useful concept based on the correlation integral is the *autocorrelation* function, which is defined as the correlation of a function with its complex conjugate (Bates and McDonnell 1986, p. 25). Thus, the autocorrelation of the function $f(\mathbf{x})$ is defined by

$$\begin{aligned} \mathcal{A}[f(\mathbf{x})] = ff(\mathbf{x}) &= f^*(\mathbf{x}) \star f(\mathbf{x}) \\ &= f(\mathbf{x}) \odot f^*(-\mathbf{x}), \end{aligned} \tag{2.26}$$

where $\mathcal{A}[\cdot]$ signifies the operation of autocorrelation and $ff(\mathbf{x})$ is the resulting autocorrelation function.

When all the component functions of a convolution are compact it follows from (2.20) that, in general, the resulting convolution is of greater extent in each coordinate direction than each of the components. This increase in the extent is, however, limited by the relationship that the extent of the convolution in each coordinate direction can be no greater than the sum of extents of each of the component functions in that coordinate direction, i.e.

$$L_{f \odot h}(x_k) \leq L_f(x_k) + L_h(x_k) \quad \text{for all} \quad k. \tag{2.27}$$

When each of the component functions is positive (§2.1.2), the equality in (2.27) holds. Following Bates and McDonnell (1986, §7), the result contained in (2.27) is called the *extent of convolution theorem.* A theorem, corresponding to that contained in (2.27) (but with $\odot$ replaced by $\star$), can be derived for correlation. The increase in

extent caused by convolution can be observed in the examples of blurring presented in Figs. 1.5 and 1.6.

If it is known that a function is positive, then the extent of that function can be uniquely deduced from knowledge of the extent of its autocorrelation. From (2.26) and (2.27) it follows that

$$L_f(x_k) = \frac{L_{ff}(x_k)}{2} \quad \text{for all} \quad k. \tag{2.28}$$

This relationship is conveniently termed the *extent of autocorrelation theorem.*

2.3 The Fourier Transform

The Fourier transform provides a unified mathematical approach to the study of a diverse range of linear systems. Electrical networks, antennas, signal analysis and image processing are just some of the topics that can usefully be studied with its aid (Bracewell 1978). Typically, one invokes the transform as a mathematical tool in order to replace a problem that cannot be solved easily with one that can be readily solved. In addition to being a powerful mathematical tool, the Fourier transform also has a definite physical meaning in many situations. Antennas, lenses, eyes and ears are just some of the artificial and naturally occurring devices whose operation is closely modelled by this transform. Table 2.2 lists further physical systems involving the Fourier transform.

Following Bates and M$^{\text{c}}$Donnell (1986, §6) (*cf.* Bracewell 1978, p. 6) the Fourier transform of the general continuous function $f(\mathbf{x})$ is defined by

$$\mathcal{F}[f(\mathbf{x})] = F(\mathbf{u}) = \int\!\!(K)\!\!\int f(\mathbf{x})\mathrm{e}^{\imath 2\pi \mathbf{u}\cdot\mathbf{x}}\,\mathrm{d}\mathbf{x}, \tag{2.29}$$

where $\mathcal{F}[\cdot]$ signifies Fourier transformation, the upper-case version of a lower case variable denotes its Fourier transform and

$$\mathbf{u}\cdot\mathbf{x} = \sum_{k=1}^{K} u_k x_k. \tag{2.30}$$

The inverse Fourier transform corresponding to (2.29) is defined by

$$\mathcal{F}^{-1}[F(\mathbf{u})] = f(\mathbf{x}) = \int\!\!(K)\!\!\int F(\mathbf{u})\mathrm{e}^{-\imath 2\pi \mathbf{x}\cdot\mathbf{u}}\,\mathrm{d}\mathbf{u}. \tag{2.31}$$

The image resulting from the Fourier transform operation is said to be the *spectrum* (Bracewell 1978) or *visibility* (Thompson *et al.* 1986) of the image. The term spectrum is adopted throughout this thesis. $f(\mathbf{x})$ and $F(\mathbf{u})$ are said to form a Fourier transform pair, denoted by $f(\mathbf{x}) \longleftrightarrow F(\mathbf{u})$. The arrow emphasizes the invertible relationship existing between the image and its transform, that is $f(\mathbf{x}) = \mathcal{F}^{-1}[F(\mathbf{u})]$, which always exists provided the integrals are convergent (Bates and M$^{\text{c}}$Donnell 1986, §6). An example of the Fourier transform of a two-dimensional image is shown in Fig. 2.4.

When $K = 2$, $\mathbf{x}$ and $\mathbf{u}$ may be replaced by the respective pairs of independent variables (x, y) and (u, v). Equation (2.29) is then written

$$F(u, v) = \iint f(x, y)\mathrm{e}^{\imath 2\pi(ux+vy)}\,\mathrm{d}x\,\mathrm{d}y. \tag{2.32}$$

Application	Object	Spectrum	Image
X-ray and neutron crystallography (Ramachandran and Srinivasan 1970)	Electron density of molecular structure	Diffraction pattern	Positive
Microscopy (acoustic, light and electron) (*cf.* Missell 1978)	Transmissivity or reflectivity of specimen (complex quantities in general)	Back focal plane field	Complex
Interferometry and spectroscopy Radio astronomical aperture synthesis Optical astronomical speckle imaging	Spatially incoherent radiating source distributions	Interferometric spatial frequency spectrum (visibility)	Positive
Radio engineering Ultrasonics Acoustics Laser optics	Coherently radiating or induced source distributions	Far-field (Fraunhofer) radiating pattern	Complex
Communications Speech processing	Signal	Temporal frequency spectrum	Real
Spectroscopy	Signal	Frequency spectrum	Real

Table 2.2: Physical systems involving the Fourier transform. (After Fright 1984.)

In polar coordinates (2.32) is expressed as (Bates and M^{c}Donnell 1986, §6)

$$F(\rho;\phi) = \mathcal{F}[f(r;\theta)] = \int_0^{\infty}\int_0^{2\pi} f(r;\theta)\mathrm{e}^{\imath 2\pi\rho r\cos(\phi-\theta)}\, r\,\mathrm{d}\theta\,\mathrm{d}r, \tag{2.33}$$

where $(r;\theta)$ (§2.1) and $(\rho;\phi)$ are the image-space and the corresponding Fourier-space polar variables. A semicolon inside parentheses signifies that the preceding and following variables are radial and angular ordinates respectively.

2.4 Properties of the Fourier Transform

Properties of the Fourier transform that are utilized in this thesis are summarized in Table 2.3. Derivations of these properties are presented in several texts (see, for example, Goodman 1968, pp. 276 ff.; Bracewell 1978; Haykin 1983, pp. 23 ff.).

Symmetry properties are important when applying Fourier theory. Therefore, terminology defining signal symmetry is introduced here in terms of a complex function $q(\mathbf{x})$. An *even* function has the property that $q(\mathbf{x}) = q(-\mathbf{x})$. An *odd* function is defined by the relationship $-q(\mathbf{x}) = q(-\mathbf{x})$. Yet another symmetry is provided when the signal $q(\mathbf{x})$ has a real part that is even and an imaginary part that is odd. Such a function is *conjugate-symmetric*, or alternatively *hermitian* (Bracewell 1978), and is succinctly described by

$$q(\mathbf{x}) = q^*(-\mathbf{x}). \tag{2.34}$$

	Property	Mathematical Description
1.	Linearity	$aq_1(\mathbf{x}) + bq_2(\mathbf{x}) \longleftrightarrow aQ_1(\mathbf{u}) + bQ_2(\mathbf{u})$
2.	Image-space shifting	$q(\mathbf{x} - \mathbf{x}_0) \longleftrightarrow Q(\mathbf{u})e^{(+i2\pi\mathbf{u}\cdot\mathbf{x}_0)}$
3.	Frequency shifting	$e^{(i2\pi\mathbf{u}_c\cdot\mathbf{x})}q(\mathbf{x}) \longleftrightarrow Q(\mathbf{u} - \mathbf{u}_c)$
4.	Area under $q(\mathbf{x})$	$\int(K)\int q(\mathbf{x})\,d\mathbf{x} = Q(0)$
5.	Area under $Q(\mathbf{u})$	$q(0) = \int(K)\int Q(\mathbf{u})\,d\mathbf{u}$
6.	Energy conservation (Parseval's Theorem)	$\int(K)\int \lvert q(\mathbf{x})\rvert^2\,d\mathbf{x} = \int(K)\int \lvert Q(\mathbf{u})\rvert^2\,d\mathbf{u}$
7.	Differentiation in image-space	$\frac{d}{d\mathbf{x}}q(\mathbf{x}) \longleftrightarrow i2\pi\mathbf{u}Q(\mathbf{u})$
8.	Conjugate functions	If $q(\mathbf{x}) \longleftrightarrow Q(\mathbf{u})$, then $q^*(\mathbf{x}) \longleftrightarrow Q^*(-\mathbf{u})$
9.	Convolution theorem	$q_1(\mathbf{x})\odot q_2(\mathbf{x}) \longleftrightarrow Q_1(\mathbf{u})Q_2(\mathbf{u})$
10.	Multiplication in image-space	$q_1(\mathbf{x})q_2(\mathbf{x}) \longleftrightarrow Q_1(\mathbf{u})\odot Q_2(\mathbf{u})$
11.	Autocorrelation	$\mathcal{A}[q(\mathbf{x})] = qq(\mathbf{x})$ $= q(\mathbf{x})\odot q^*(-\mathbf{x}) \longleftrightarrow \lvert Q(\mathbf{u})\rvert^2$
12.	Reality in image-space	If $\mathcal{I}[q(\mathbf{x})] = 0$, then $Q(\mathbf{u}) = Q^*(-\mathbf{u})$
13.	Even symmetry in image-space	If $q(\mathbf{x}) = q(-\mathbf{x})$, then $Q(\mathbf{u}) = Q(-\mathbf{u})$
14.	Conjugate symmetry in image-space	If $q(\mathbf{x}) = q^*(-\mathbf{x})$, then $\mathcal{I}[Q(\mathbf{u})] = 0$

Table 2.3: A summary of properties of the Fourier transform.

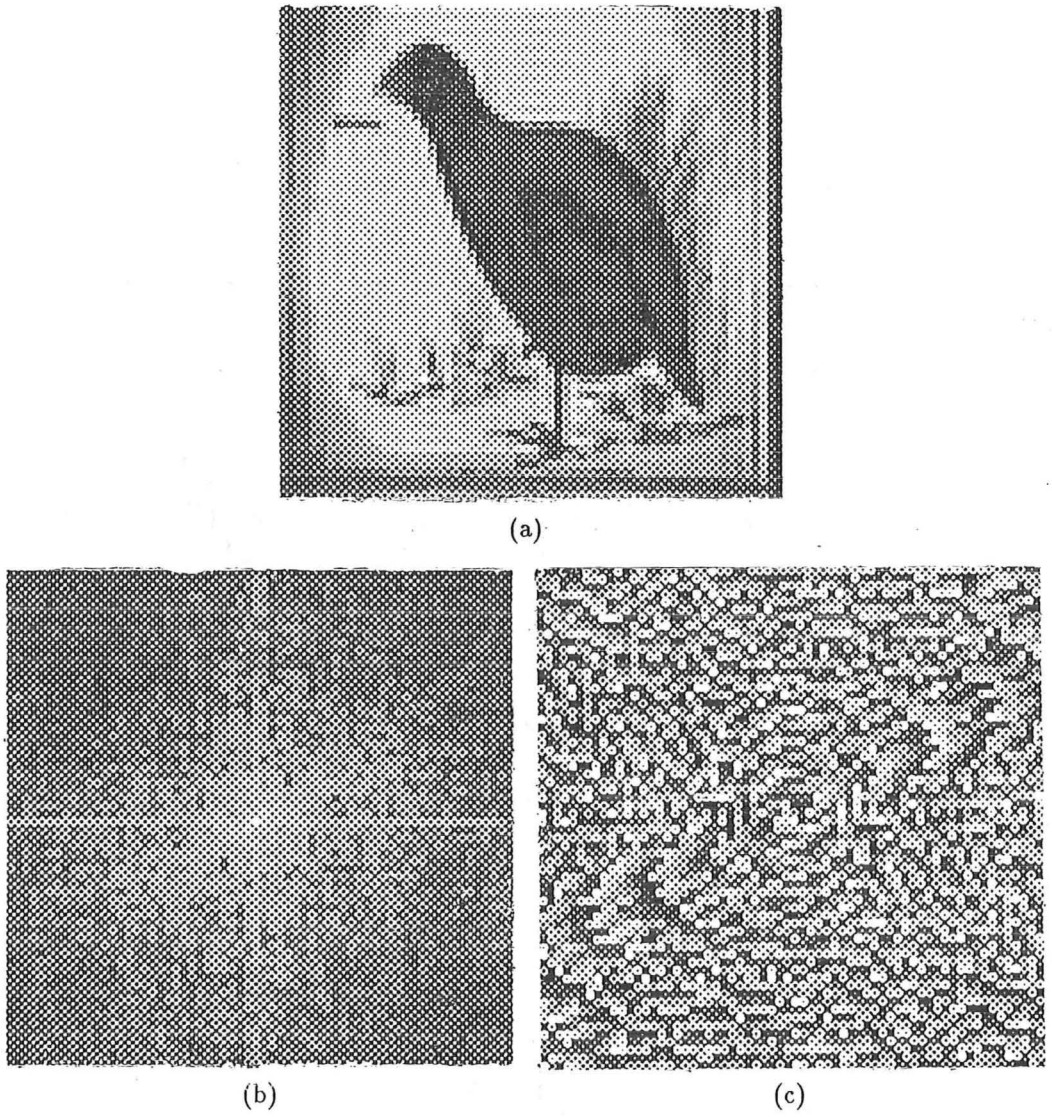


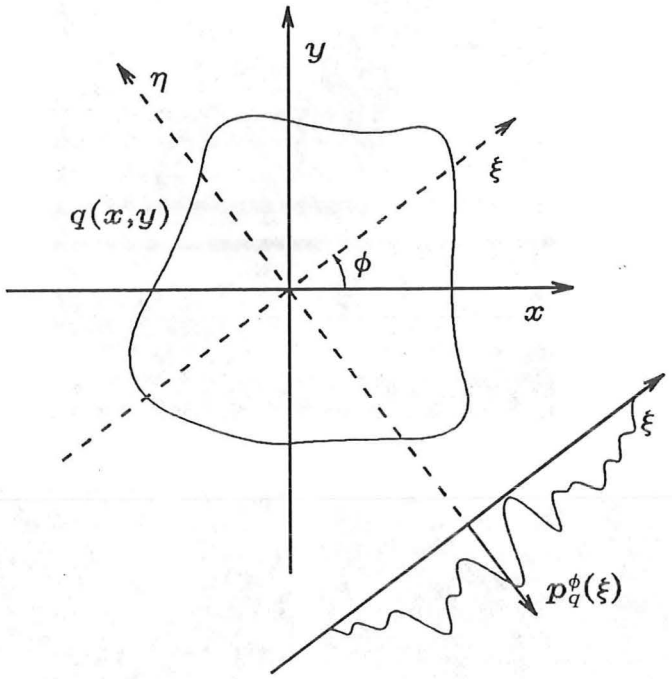
Figure 2.4: Example of the Fourier transform of a two-dimensional image: (a) $f(\mathbf{x})$; (b) $\ln|F(\mathbf{u})|$; (c) $\mathcal{P}[F(\mathbf{u})]$.

A further useful property can be derived by combining properties 12 and 13 from Table 2.3. The spectrum of a function that is both real and even-symmetric is itself real and symmetric. This property has important practical implications when removing certain classes of symmetric blurring, for example that caused by an out-of-focus camera (see §1.4).

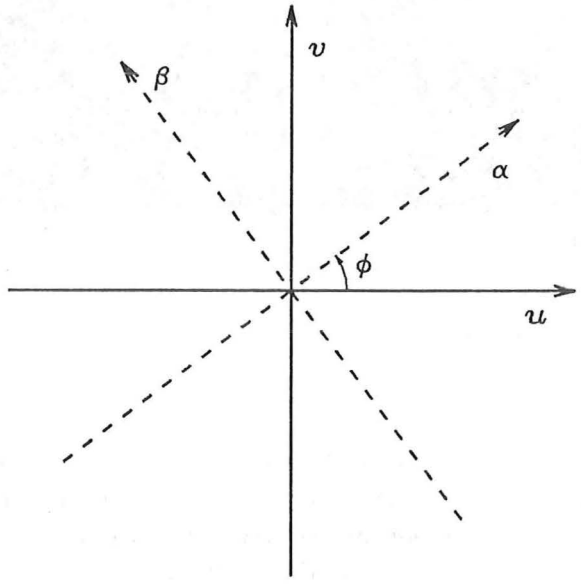
A complete list of symmetry properties of the Fourier transform is presented by Bracewell (1978, pp. 14 ff.).

2.4.1 The Projection Theorem

The projection theorem (Garden 1984; Bates and McDonnell 1986, §9) is a property of the Fourier transform that is applied extensively in several fields of image processing (Herman 1979), notably comput(eris)ed tomography (CT) (Bates *et al.* 1983; Garden 1984) and multi-element interferometry in radio astronomy (Napier *et al.* 1983). This theorem is now discussed with reference to Fig. 2.5.



(a)



(b)

Figure 2.5: The projection theorem in two dimensions: (a) the two-dimensional image $q(x, y)$ is projected, or integrated, in the direction of the η-axis for each value of ξ to produce the one-dimensional image $p_q^\phi(\xi)$. This image is called the projection of $q(x, y)$ at angle ϕ (refer to (2.35)); (b) Fourier-space corresponding to (a). The projection theorem reveals that $S_Q^\phi(\alpha)$, the slice through Fourier-space along the α-axis, is obtained from the one-dimensional Fourier transform of $p_q^\phi(\xi)$, i.e. $S_Q^\phi(\alpha) \longleftrightarrow p_q^\phi(\xi)$. This corresponds to the values of $Q(u, v)$ along the line inclined at angle ϕ to the u-axis.

Projections are conveniently defined in terms of two pairs of Cartesian coordinates (x, y) and (ξ, η) having a common origin, but with (ξ, η) rotated in image-space through an arbitrary angle ϕ (see Fig. 2.5). Suppose $q(x, y)$ is the two-dimensional image whose projection is to be computed. The projection in the η-direction (often called the projection at angle ϕ) is formed by calculating the line integral of $q(x, y)$ in the η-direction for all ξ. On denoting this projection by $p_q^\phi(\xi)$ (where the superscript ϕ indicates the dependence of the projection upon angle), it can be seen that

$$p_q^\phi(\xi) = \int_{-\infty}^{\infty} q(x, y)\, \mathrm{d}\eta. \tag{2.35}$$

$p_q^\phi(\xi)$ is also termed the *Radon transform* of $q(x, y)$ (Rosenfeld and Kak 1982). Parallel projections are implicit in the use of the term projection in this thesis, although other types of projection, notably fan-beam projections, arise in some practical imaging situations (Rosenfeld and Kak 1982; Garden 1984).

Consider the Cartesian coordinates (α, β) in two-dimensional Fourier-space, rotated by ϕ (the same amount as the original projection) with respect to the Cartesian coordinates (u, v) (see Fig. 2.5(b)). It follows that

$$ux + vy = \alpha\xi + \beta\eta, \tag{2.36}$$

and

$$u = \alpha\cos(\phi) - \beta\sin(\phi) \quad \text{and} \quad v = \alpha\sin(\phi) + \beta\cos(\phi). \tag{2.37}$$

Remembering that, by definition, $q(x, y) \longleftrightarrow Q(u, v)$, one can straightforwardly deduce that $q(\xi, \eta) \longleftrightarrow Q(\alpha, \beta)$, since the orientation of the original Cartesian coordinates, i.e. (x, y) and (u, v), could equally validly have been chosen to be that of (ζ, γ) and (α, β). Consider now the values of $Q(u, v)$ along the line rotated by ϕ from the u-axis. This corresponds to $Q(\alpha, 0)$ and is called a *slice* of $Q(u, v)$ at angle ϕ. It is here denoted $S_Q^\phi(\alpha)$, where the superscript ϕ distinguishes it from the support of the image (see §2.1.4). Combining (2.32), (2.35) and (2.36) this slice may be written

$$\begin{aligned} S_Q^\phi(\alpha) &= Q(\alpha, 0) \\ &= \iint q(\xi, \eta) \mathrm{e}^{\imath 2\pi\xi\alpha}\, \mathrm{d}\xi\, \mathrm{d}\eta \\ &= \int \left[\int q(\xi, \eta)\, \mathrm{d}\eta \right] \mathrm{e}^{\imath 2\pi\xi\alpha}\, \mathrm{d}\xi \\ &= \int p_q^\phi(\xi) \mathrm{e}^{\imath 2\pi\xi\alpha}\, \mathrm{d}\xi \\ &= \mathcal{F}\left[p_q^\phi(\xi) \right]. \end{aligned} \tag{2.38}$$

Thus, the one-dimensional Fourier transform of the projection $p_q^\phi(\xi)$ provides the slice of the two-dimensional Fourier transform $Q(u, v)$ corresponding to $\beta = 0$. This relationship is called the *projection theorem.*

Consider the convolution $g(x, y) = f(x, y) \circledcirc h(x, y)$. The theory of image reconstruction from projections (Bates and McDonnell 1986, Chapter 5) ensures that the projection $p_g^\phi(\xi)$, of $g(x, y)$, is the one-dimensional convolution of the projections, $p_f^\phi(\xi)$ and $p_h^\phi(\xi)$, of $f(x, y)$ and $h(x, y)$ respectively. Thus, the projection operation (2.35) preserves the operation of convolution even though the dimensionality is reduced.

2.5 Implementing the Fourier Transform

To obtain the Fourier transform of a sampled image, it is convenient to invoke an algorithm called the *discrete Fourier transform* (DFT) (Bracewell 1978, Chapter 18). This algorithm relates point samples of a compact image to point samples of its Fourier transform. The sample points are required to lie on a uniformly spaced rectangular grid (Bates and McDonnell 1986, §12). Let $q(x)$ be the one-dimensional sampled function

$$q(x) = \sum_{n=0}^{N-1} q_n \delta(x - x_n), \tag{2.39}$$

where N is the number of image-space samples, q_n are the sample values and

$$x_n = n\Delta \tag{2.40}$$

are the sample positions. It is assumed here that the image is of finite extent so N is finite.

Combining (2.39) and (2.29) yields the DFT, which shows that the Fourier-space samples Q_m, corresponding to the image-space samples q_n, can be written

$$Q_m = \sum_{n=0}^{N-1} q_n e^{\imath 2\pi m(\frac{n}{N})}, \tag{2.41}$$

where $Q_m = Q(m/N)$ and $q_n = q(n/N)$. N is both the number of image-space samples and the number of Fourier-space samples. The corresponding *inverse discrete Fourier transform* is written as

$$q_n = \frac{1}{N} \sum_{n=0}^{N-1} Q_m e^{-\imath 2\pi n(\frac{m}{N})}. \tag{2.42}$$

The DFT can be straightforwardly extended to two-or-more dimensions (Bracewell 1978).

The process of sampling an image causes its spectrum to be periodic, with a period $1/\Delta$, as shown in Fig. 2.6. If the image is not *band-limited* to the Nyquist frequency (i.e. non-zero data in its spectrum are not confined to the range of frequencies $-1/2\Delta$ to $1/2\Delta$), then aliasing occurs. If, in addition, the Fourier transform is sampled, the image is also assumed to be periodic, with a period $N\Delta$. Thus, implicit in the use of the DFT, where both image- and Fourier-space are sampled, is the idea that the data in each space are repetitive (see Fig. 2.6(c)).

In 1965, a computationally efficient algorithm to implement the DFT became widely known (Cooley and Tukey 1965). This implementation, called the *fast Fourier transform* (FFT) algorithm (Ramirez 1985), is now almost universally used to perform Fourier transform operations in a digital computer. By reformulating the DFT to minimize the number of multiplications required, a considerable speed increase is achieved. This speed advantage is revealed by considering the Fourier transform of a one-dimensional function. In this situation, it can be deduced from (2.41) that, if a direct algorithmic translation of this equation is invoked, N^2 multiplications are required to calculate the DFT. However, the number of multiplications to implement an FFT is of the order of $N \log_2 N$ (Bracewell 1978, p. 371), an extremely significant improvement for applications where N is large. When generating computational results for this thesis, all Fourier transforms were computed by invoking an FFT routine. This is appropriate since many of the algorithms comprise iterative loops involving Fourier transforms of multidimensional data at each iteration.

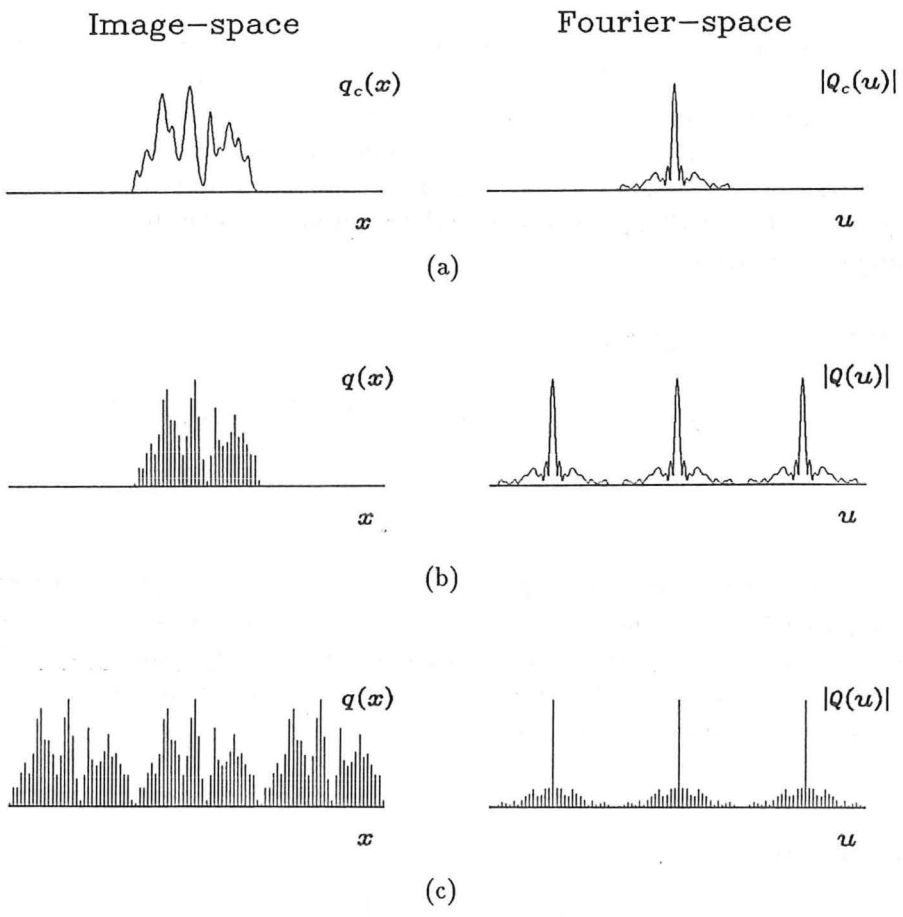


Figure 2.6: The result of sampling on the Fourier transform operation. (a) the continuous function $q_c(\mathbf{x})$ and the magnitude of its Fourier transform $|Q_c(u)|$; (b) sampling the function causes the Fourier transform to be repetitive; (c) sampling the Fourier transform in addition to the function causes the function to be repetitive.

Discussion in this thesis regarding the implementation of the Fourier transform has so far been limited to the use of conventional digital computers. It should be noted however that an optical implementation would, if realizable, be many orders of magnitude faster. A distinction must be drawn here between digital and analogue optical computers. Much research is currently being directed towards the incorporation of optical elements within a conventional digital computer (Arrathoon 1986). By performing certain critical functions much faster than their semiconductor equivalents, optical switching components increase the overall speed of the computer. In terms of implementing operations such as Fourier transforms, this clearly can be considered simply as an increase in the speed of a digital computer.

However, the analogue type of computer, which would be implemented using an optical bench (Goodman 1968), is a fundamentally different concept (Andrews and Hunt 1977, §1.1). Incredibly fast processing is possible in principle, when applying this analogue-type processing because of the massive parallelism inherent in its use. For example, a simple lens 'computes' the Fourier transform of a one- or two-dimensional function in just a few picoseconds (the time for light to propagate through the lens) (Goodman 1968). There are no great difficulties, or large expense, involved in realizing a Fourier transform lens to manipulate coherent laser light (*cf.* Bates 1982b, §5.3). However, problems of accurately inputing and re-

trieving the signal from the system limit the use of such computational techniques. Glaser (1987) has recently reviewed details of optical processing with spatially incoherent light. He describes possible schemes for representing bipolar-real and complex functions and the implementation of various transforms with optical systems. Initial implementations of practical imaging algorithms using analog optical processing have been reported (*cf.* Peri 1987), but the digital computer continues to dominate image processing computation.

2.6 The z-transform

It is often useful to treat the DFT as a polynomial by rewriting (2.41) as

$$Q(u) = \sum_{n=0}^{N-1} q_n e^{\imath n u \Omega}, \tag{2.43}$$

where Ω is called the *fundamental frequency* (Requicha 1980) and incorporates all scaling required to relate the image samples to the Fourier-space samples. The q_n are called the *coefficients* of the polynomial, and are the image samples (see (2.39)). The quantity $N-1$ is called the *order*, or *degree*, of the polynomial.

Upon defining the transformation

$$\zeta = e^{\imath u \Omega}, \tag{2.44}$$

(2.43) can be rewritten as

$$Q(\zeta) = \sum_{n=0}^{N-1} q_n \zeta^n, \tag{2.45}$$

which is the discrete z-transform (Stanley 1975, Chapter 4). $Q(\zeta)$, the z-transform of $q(x)$, is a polynomial of degree $N-1$ in the complex variable ζ. Invoking the notation $\mathcal{Z}[\cdot]$ to denote the transformation from image-space to z-space, (2.45) can be written $Q(\zeta) = \mathcal{Z}[q(x)]$, and the corresponding inverse transformation can be written $q(x) = \mathcal{Z}^{-1}[Q(\zeta)]$.

The convolution theorem (see Table 2.3) also holds for the z-transform of a convolution (Oppenheim and Schafer 1975, §2.3.8). Thus, the z-transform of the N-composite convolution, (2.24), is

$$G(\zeta) = F_1(\zeta)F_2(\zeta)\ldots F_N(\zeta). \tag{2.46}$$

Since each of the $F_n(\zeta)$ are polynomials, it follows that convolution in image-space corresponds to the multiplication of polynomials in z-space. The N-polynomials are the z-transforms of the components of the convolution.

It is appropriate to briefly consider the question of deconvolution here. Since, from (2.46), the z-transform of an image-space convolution is the product of the N polynomials comprising the N components' z-transforms, it follows that deconvolution in image-space is equivalent to polynomial factorization in z-space. Thus, if one is able to factorize the z-transform of a function, one is effectively perfoming image-space deconvolution.

One-dimensional polynomials, i.e. polynomials of a single variable, can always be factored and are therefore said to be *reducible*. Consequently it is always possible to find two (or more) one-dimensional functions, which when convolved together, produce the given convolution. This is the well known fundamental theorem of algebra (Kreysig 1979, p. 653). However, no such theorem exists for polynomials of

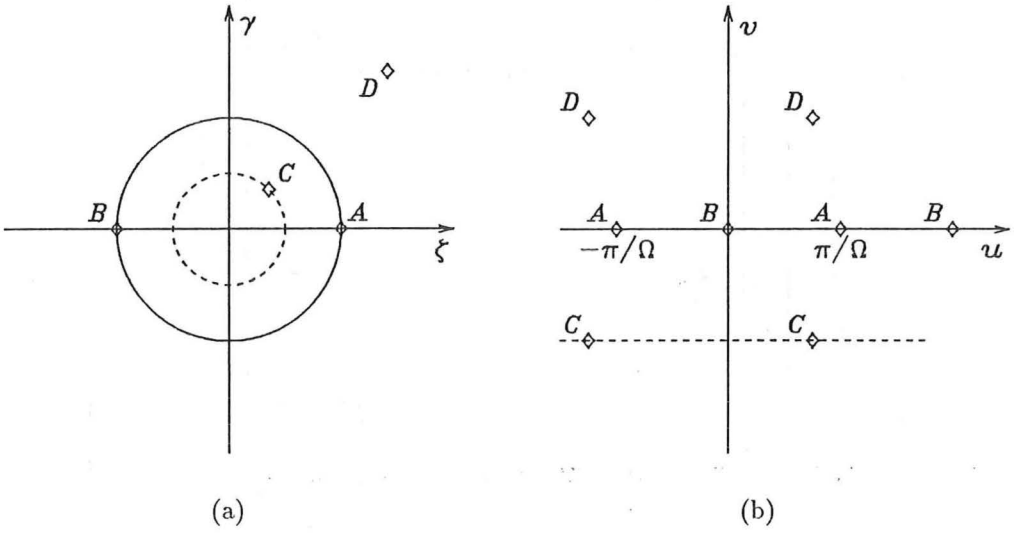


Figure 2.7: A conformal mapping exists between z-space (a) and complex Fourier-space (b). The unit circle in z-space maps to the real axis of complex Fourier-space. A strip of width $2\pi/\Omega$ in Fourier-space maps to the entire z-space.

more-than-one variable (Hayes and M$^{\text{c}}$Clellan 1982). Thus, deconvolution of functions having more-than-one dimensions is inherently different than deconvolution of one-dimensional functions (Bruck and Sodin 1979; Lane 1988). The consequences of this argument are developed further in §§5.3 and 7.1.

2.7 Analytic Continuation

It is often useful to consider functions of a complex variable, although, in practice, measurements are usually made as a function of a real variable. Such measurements can easily be transformed to become a function of a complex variable simply by allowing the real variable to become complex. This procedure is a form of analytic continuation (Kreysig 1979, p. 679).

It is useful to analytically continue the DFT by generalizing the real variable u in (2.41) to the complex variable

$$w = u + \imath v, \tag{2.47}$$

where v, the imaginary part of w, is not to be confused with the Cartesian coordinate in Fourier-space (refer to Table 2.1). Equation (2.43) then becomes

$$F(w) = \sum_{n=0}^{N-1} f_n \mathrm{e}^{\imath n w \Omega} \tag{2.48}$$

and the transformation (2.44) can be viewed as a conformal mapping between z-space and complex Fourier-space. A vertical strip of width $2\pi/\Omega$ in complex Fourier-space is mapped to the entire z-space, as shown in Fig. 2.7. The relationship between image-space, Fourier-space and z-space is illustrated in Fig. 2.8.

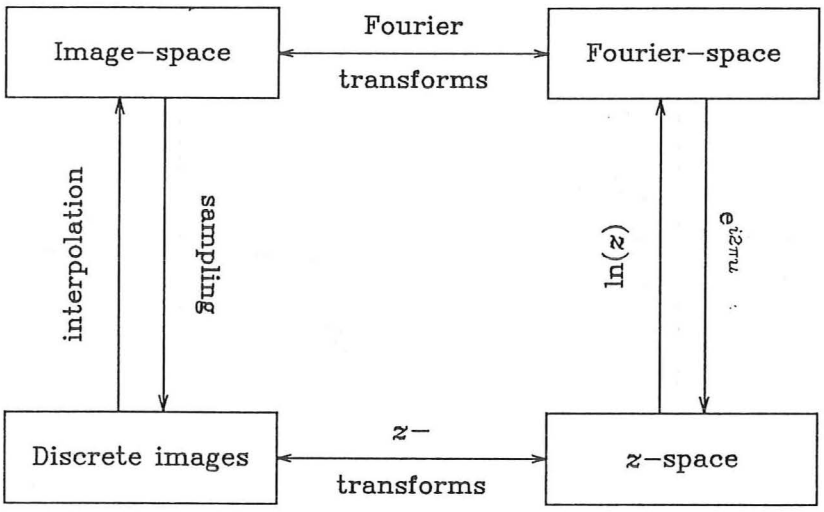


Figure 2.8: The relationships between Fourier-space, z-space and image-space (after Lane 1988).

2.8 Entire Functions

Entire functions are introduced in this section since properties of these functions are fundamental to many of the algorithms presented in Chapters 5, 6 and 7.

A function is *analytic* at a point if it is both defined and differentiable, i.e. it is 'well behaved', everywhere in the neighbourhood of the point (Kreysig 1979, p 578). Alternatively, a function is said to be analytic at a point if it can be represented, in the neighbourhood of the point, by a Taylor series expansion (Freiberger 1960, p. 39). A function is *entire* if it is analytic at all points in a finite complex plane (Requicha 1980, p. 312).

Entire functions of a complex variable have found widespread application in modelling of physical phenomena (see, for example, Bates 1969; Nakajima and Asakura 1983; Stefanescu 1985) because, when one member of a Fourier transform pair has finite support, the other can be modelled as an entire function (Requicha 1980, p. 308). For example, a signal that is band-limited can be considered to be an entire function because, by definition, its Fourier transform has finite support.

An *entire function of exponential type* (EFET) is an entire function whose behaviour as $|w|$ tends to ∞ is of the order of $e^{C|z|}$ where C is a constant (Titchmarsh 1932, Chapter 8). EFETs are the most important class of entire functions for engineering applications (Requicha 1980), largely because algebraic and trigonometric polynomials are members of this class of functions.

An important property of EFETs is that they are completely characterized, apart from a complex scaling constant, by their zeros of which there are a denumerable infinity (Levin 1964). Thus, knowledge of the zeros of the function is equivalent to knowledge of the function itself. Napier (1971, p. 14) notes that, although the concept of representing $F(u)$ in terms of its zero positions is less common than the usual samples, it often provides a more useful representation.

Chapter 3

An Introduction to Deconvolution

Whereas Chapter 2 introduces general image processing and mathematical terminology and concepts, this chapter provides an introduction to deconvolution techniques and thereby delves more specifically into the topic that is central to this thesis. The discussion contained here continues laying the groundwork, begun in Chapters 1 and 2, for the algorithms discussed in subsequent chapters.

Descriptions of several established algorithms for implementing the first deconvolution problem posed in §1.5, conventional deconvolution, are given in §3.1. The conventional deconvolution algorithms presented in this section are some of those that are commonly applied in the image and signal processing applications discussed in this thesis. Many other algorithms find application in these and various other image (for example, Huang 1981), and signal (*cf.* Schafer *et al.* 1981), processing applications.

An algorithm implementing ensemble blind deconvolution is presented in §3.2. This algorithm is due to Stockham, Jr. (1971) and has been applied to restore degraded music recordings. The technique is also capable of restoring limited classes of blurring of images. Chapters 4 and 6 are both devoted to describing other algorithms that implement ensemble blind deconvolution.

All real-world measurements are subject to noise, whether that noise is induced by limitations of the recording device, or the necessity to make remote measurements, or any other cause. Thus, this chapter concludes with a brief discussion of problems inherent in performing deconvolution in the presence of contamination (§3.3).

It should be remembered that an isoplanatically blurred image $g(\mathbf{x})$ can be expressed as (see (1.4))

$$g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x}) + c(\mathbf{x}), \tag{3.1}$$

where $f(\mathbf{x})$ is the true object, $h(\mathbf{x})$ is the psf causing the blurring, and $c(\mathbf{x})$, the contamination, incorporates any noise or other departures from the convolutional model. The algorithms discussed in this chapter are described in terms of this notation.

3.1 Conventional Deconvolution

Following the definition introduced in §1.5 the conventional deconvolution problem is posed as:

> "Given $g(\mathbf{x})$ and an estimate $\hat{h}(\mathbf{x})$, of $h(\mathbf{x})$, retrieve $\hat{f}(\mathbf{x})$, an estimate of $f(\mathbf{x})$."

The simplest conventional deconvolution algorithm, inverse filtering, is introduced in §3.1.1. The *inverse filter* is a multiplicative filter that is applied in Fourier-

space. Unfortunately, it tends to produce severe artefacts in the presence of contamination, a tendency which the *Wiener filter* (Bates and McDonnell 1986) mostly overcomes. This latter filter, which is described in §3.1.2, is a modification of the inverse filter that produces a reasonably faithful estimate of the object even in the presence of appreciable contamination. Homomorphic deconvolution (Oppenheim *et al.* 1968), which is introduced in §3.1.3, reduces the convolution to an addition by forming the complex logarithm of the spectrum of $g(\mathbf{x})$. Removal of the blurring then involves subtracting the estimate of the psf after it has been subjected to similar processing. Although the deconvolution is implemented as a subtraction, the homomorphic filter is effectively equivalent to the inverse filter, and therefore suffers from the same sensitivity to contamination.

The iterative subtractive deconvolution algorithm, CLEAN, is the subject of §3.1.4. In contrast to the inverse, Wiener and homomorphic filters, CLEAN operates in image-space and can be readily adapted to point spread variant blurring (Bates *et al.* 1982).

3.1.1 Inverse Filtering

The convolution theorem (§2.4) reveals the rationale behind the most simple conventional deconvolution procedure, the *inverse filter*. Consider the blurred image $g(\mathbf{x})$, as expressed by (3.1), for the special case where $c(\mathbf{x})$ is negligible. The convolution theorem (see Table 2.3) confirms that the Fourier transform, $G(\mathbf{u})$, is the product of $F(\mathbf{u})$ and $H(\mathbf{u})$ i.e.

$$G(\mathbf{u}) = F(\mathbf{u})H(\mathbf{u}). \tag{3.2}$$

If some estimate $\hat{H}(\mathbf{u}) \longleftrightarrow \hat{h}(\mathbf{x})$ of the psf is provided then a filter

$$\Omega(\mathbf{u}) = \frac{1}{\hat{H}(\mathbf{u})} \tag{3.3}$$

may be defined. The product of this filter and $G(\mathbf{u})$ yields an estimate $\hat{F}(\mathbf{u})$, of $F(\mathbf{u})$, i.e.

$$\begin{aligned} \hat{F}(\mathbf{u}) &= G(\mathbf{u})\Omega(\mathbf{u}) \\ &= \frac{F(\mathbf{u})H(\mathbf{u})}{\hat{H}(\mathbf{u})} \\ &\approx F(\mathbf{u}), \end{aligned} \tag{3.4}$$

provided $\hat{H}(\mathbf{u})$ is an adequately faithful estimate of $H(\mathbf{u})$. The filter $\Omega(\mathbf{u})$ is known as the inverse filter. An estimate of the object is obtained by inverse Fourier transforming the LHS of (3.4), i.e. $\hat{f}(\mathbf{x}) \longleftrightarrow \hat{F}(\mathbf{u})$.

The inverse filter takes account of neither the inevitable contamination which appears in a practical imaging situation, nor the discrepancies between the actual psf and the available estimate of it. When contamination is present the reconstructed estimate is given by

$$\hat{F}(\mathbf{u}) = F(\mathbf{u}) + \frac{C(\mathbf{u})}{\hat{H}(\mathbf{u})}. \tag{3.5}$$

Since it appears in the denominator of the second term on the RHS of (3.5), wherever $|\hat{H}(\mathbf{u})|$ is small the effect of the contamination is amplified, leading to a degraded reconstruction. The degradation is often manifest as ringing in the reconstruction at frequencies corresponding to those for which $|\hat{H}(\mathbf{u})| \ll |\hat{C}(\mathbf{u})|$.

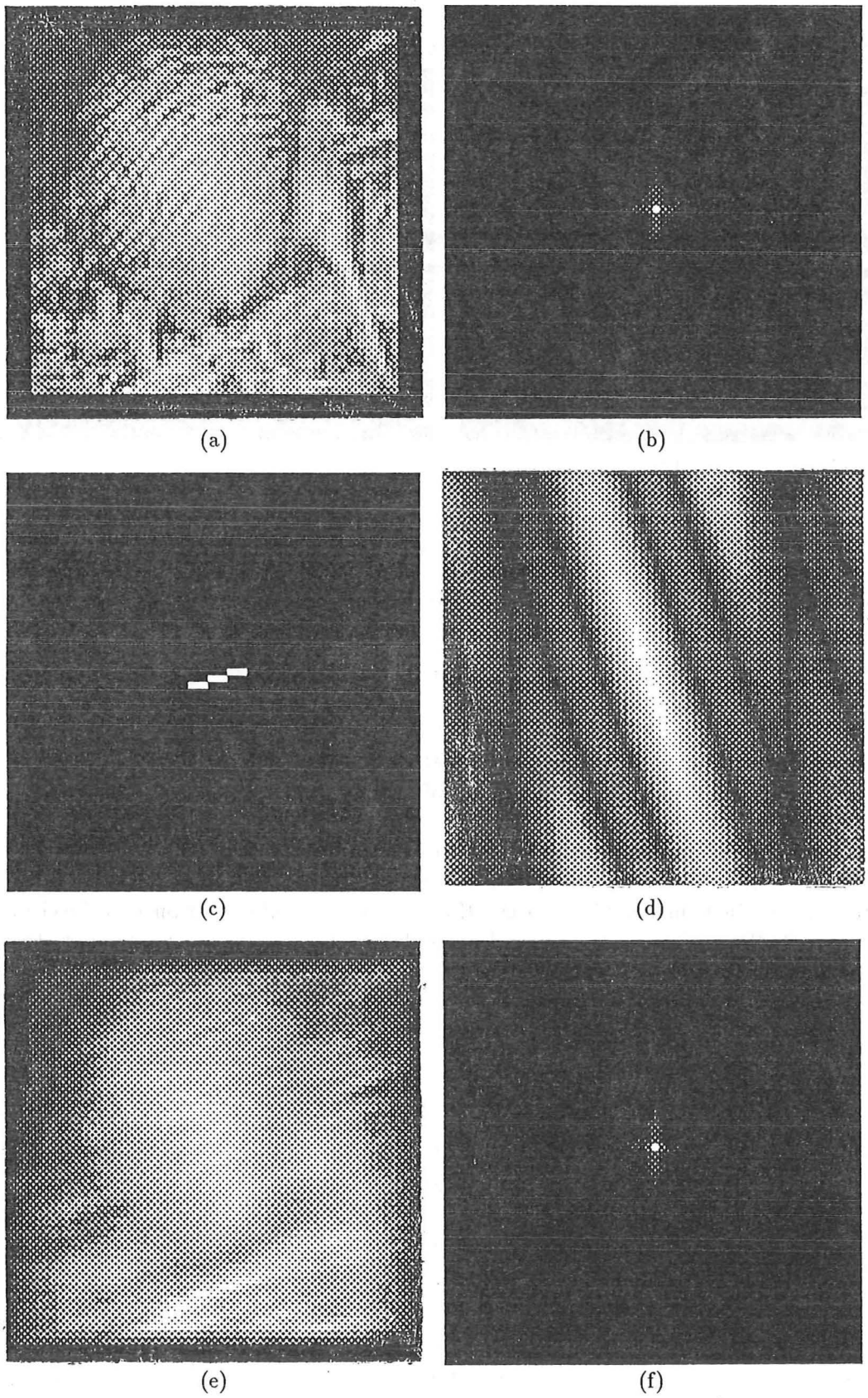
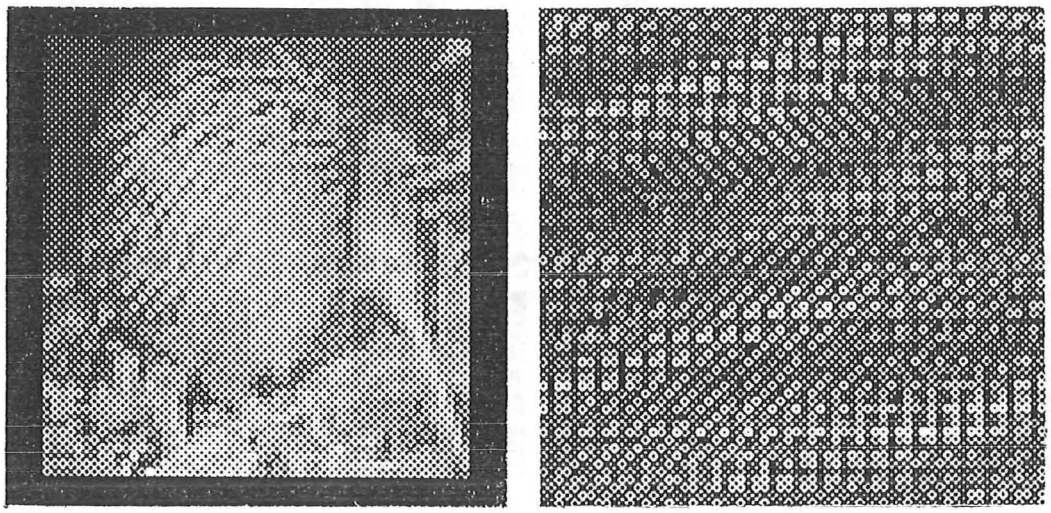


Figure 3.1: Generation of the convolution used for Figs. 3.2 and 3.4 (*cf.* Fig. 1.5): (a) the true object $f(\mathbf{x})$; (b) $|F(\mathbf{u})|$; (c) the psf, $h(\mathbf{x})$, typical of that produced by moving a camera when making an exposure; (d) $|H(\mathbf{u})|$; (e) the blurred image $g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x})$; (f) $|G(\mathbf{u})| = |F(\mathbf{u})H(\mathbf{u})|$.



(a) (b)

Figure 3.2: $\hat{f}(\mathbf{x})$ obtained by inverse filtering Fig. 3.1(e) with Fig. 3.1(c): (a) with no contamination added to $g(\mathbf{x})$; (b) when $\epsilon_{\mathrm{dB}} = -30\mathrm{dB}$ contamination is added to $g(\mathbf{x})$ before the filtering. Note that $\Omega(\mathbf{u})$ is identical for the generation of both reconstructions.

One way in which a convolution is often generated in the real world is illustrated in Fig. 3.1. When the true object shown in Fig. 3.1(a) is photographed by a moving camera, whose movement is characterized by the psf depicted in Fig. 3.1(c), the exposure that results is shown in Fig. 3.1(e). Note that the blurred image shown in Fig. 3.1(e) is not degraded by any species of contamination.

Reconstructions of the blurred image shown in Fig. 3.1(e) with an inverse filter derived from the psf depicted in Fig. 3.1(c) are presented in Fig. 3.2. The reconstruction from the uncontaminated convolution (i.e. Fig. 3.2(a)) compares favourably with the original image, Fig. 3.1(a). However, when contamination to a level of $\epsilon_{\mathrm{dB}} = -30\mathrm{dB}$ is added to the blurred image before the inverse filter is applied, the resulting reconstruction (Fig. 3.2(b)) reveals the lack of robustness of the inverse filter in the presence of contamination.

3.1.2 Wiener Filtering

A modification of the inverse filter, called the *Wiener filter*, reduces the noise amplification property of the inverse filter. The Wiener filter, $W(\mathbf{u})$, is defined (Bates *et al.* 1984; Bates and McDonnell 1986, §16) by

$$W(\mathbf{u}) = \frac{\hat{H}^*(\mathbf{u})}{\left(|\hat{H}(\mathbf{u})|^2 + \Phi(\mathbf{u})\right)}, \tag{3.6}$$

and

$$\hat{F}(\mathbf{u}) = G(\mathbf{u})W(\mathbf{u}), \tag{3.7}$$

where the function $\Phi(\mathbf{u})$ is a measure of the noise-to-signal energy ratio which is a function of frequency in general. In many practical applications the high spatial frequencies tend to be 'noisier' than their low frequency counterparts. In order to compensate for this effect it is useful to choose a $\Phi(\mathbf{u})$ such that $|\Phi(\mathbf{u})|$ increases with increasing $|\mathbf{u}|$. Such a function is illustrated in Fig. 3.3.

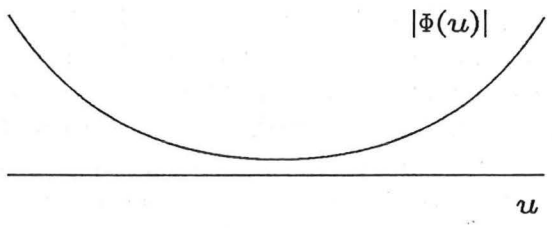
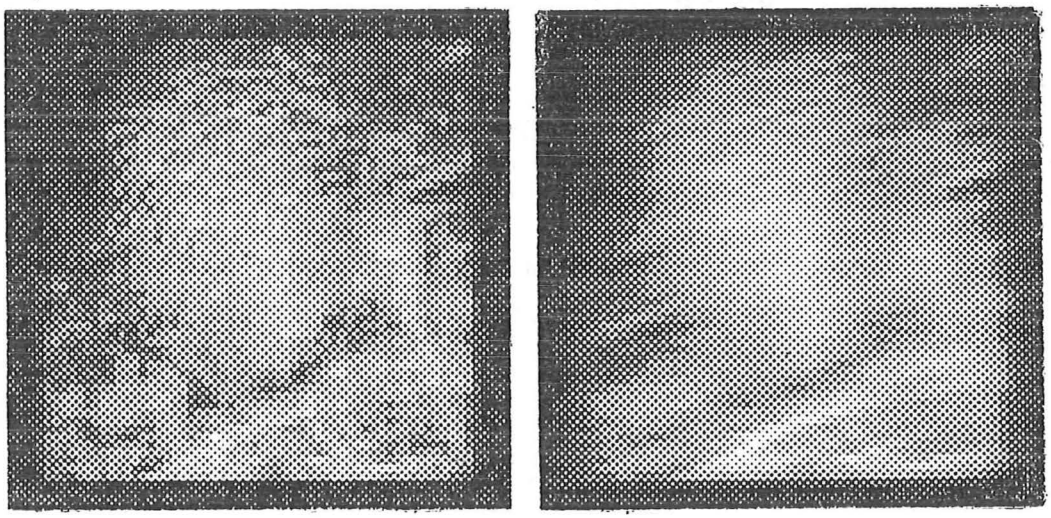


Figure 3.3: The magnitude of a typical Wiener filter function, (i.e. $|\Phi(u)|$), for a one-dimensional conventional deconvolution situation in which it is known that the contamination increases in severity with increasing frequency.



(a) (b)

Figure 3.4: $\hat{f}(\mathbf{x})$ obtained by filtering Fig. 3.1(e) (with a level of contamination specified by $\epsilon_{\text{dB}} = -30\text{dB}$), with a Wiener filter derived from Fig. 3.1(c) with: (a) an optimum value of Φ; (b) with Φ too large.

In practice it is usually difficult to accurately estimate the detailed form of $\Phi(\mathbf{u})$ and often white noise is assumed. Under this assumption $\Phi(\mathbf{u})$ is replaced with a constant Φ (Bates and M$^{\text{c}}$Donnell 1986, §16). The filter described by (3.6) then becomes

$$W(\mathbf{u}) = \frac{\hat{H}^*(\mathbf{u})}{\left(|\hat{H}(\mathbf{u})|^2 + \Phi\right)}. \tag{3.8}$$

If knowledge of the noise-to-signal energy is available, an informed choice of the value of Φ can be made. Often such information is not available, so the filter operation is performed several times, each time with a different value of Φ. The value that results in the most satisfactory result (according to whatever criterion is deemed appropriate), is chosen as the best estimate of the noise-to-signal ratio (Bates and M$^{\text{c}}$Donnell 1986, §16).

An example of the efficacy of a Wiener filter, operating under the assumption of white noise, is shown in Fig. 3.4. The convolution, Fig. 3.1(e), with contamination added to a level of $\epsilon_{\text{dB}} = -30\text{dB}$, is deconvolved with the aid of a Wiener filter

derived from the uncontaminated psf shown in Fig. 3.1(c). Comparing Fig. 3.2(b) and Fig. 3.4(a) reveals that the Wiener filter does not suffer from the noise amplification property of the inverse filter.

Note that, in the absence of noise (i.e. $H(\mathbf{u}) \gg C(\mathbf{u})$), Φ is best chosen to be zero, in which case the Wiener filter effectively reduces to the inverse filter (3.3). Note further that Φ in the denominator of RHS (3.8) adds a bias to the deconvolution operation. Since $|H(\mathbf{u})|$ tends to reduce with increasing spatial frequency, it tends to be the high spatial frequency components that are most severely affected. This is apparent in Fig. 3.4(b) where the Wiener filter constant has been chosen to be too large. Here, the high frequency components have been much more severely degraded leading to a reconstruction in which sharp edges and lines (i.e. the high frequency information) have been lost. This point is discussed further in §3.3.

3.1.3 Homomorphic deconvolution

Homomorphic deconvolution (Oppenheim *et al.* 1968; Oppenheim and Schafer 1975, Chapter 10) is another conventional deconvolution technique which, like the inverse and Wiener filters, effectively operates in Fourier-space. However, in homomorphic deconvolution, a convolution is reduced to an addition by forming the complex logarithm of the convolution's spectrum. For convenience, it is assumed that $C(\mathbf{u}) = 0$ in the following discussion.

The complex logarithm of the spectrum of $g(\mathbf{x})$, i.e.

$$\ln[G(\mathbf{u})] = \ln[F(\mathbf{u})] + \ln[H(\mathbf{u})], \tag{3.9}$$

is often called the *cepstrum* of $g(\mathbf{x})$ (Oppenheim and Schafer 1975, p. 500). When $\hat{h}(\mathbf{x})$ is known *a priori*, $\ln\left[\hat{H}(\mathbf{u})\right]$ can be evaluated and subtracted from $\ln[G(\mathbf{u})]$ leaving an estimate of the cepstrum of $f(\mathbf{x})$ i.e.

$$\begin{aligned} \ln[G(\mathbf{u})] - \ln\left[\hat{H}(\mathbf{u})\right] &= \ln[F(\mathbf{u})] + \ln[H(\mathbf{u})] - \ln\left[\hat{H}(\mathbf{u})\right] \\ &\approx \ln\left[\hat{F}(\mathbf{u})\right]. \end{aligned} \tag{3.10}$$

Provided $\hat{h}(\mathbf{x})$ is a sufficiently faithful estimate of $h(\mathbf{x})$,

$$\hat{f}(\mathbf{x}) = \mathcal{F}^{-1}\left[e^{\ln[\hat{F}(\mathbf{u})]}\right] \tag{3.11}$$

is a faithful estimate of $f(\mathbf{x})$. Homomorphic deconvolution is equivalent to inverse filtering and is therefore subject to the same limitations in the presence of contamination. However, a homomorphic Wiener filter (Bates and M^{c}Donnell 1986, §16)

$$W_h(\mathbf{u}) = \ln[W(\mathbf{u})], \tag{3.12}$$

where $W(\mathbf{u})$ is defined by (3.6), can be usefully introduced. This is essentially equivalent to the Wiener filter (§3.1.2), and ameliorates the noise amplification property of the simple homomorphic filter $\ln\left[\hat{H}(\mathbf{u})\right]$.

3.1.4 CLEAN

An alternative to the deconvolutional methods described in §§3.1.1–3.1.3 is the CLEAN algorithm (Högbom 1974), which was originally developed for image restoration in a radio astronomy context. CLEAN has, however, also found application in general image processing (Bates *et al.* 1984) and other scientific fields. For example,

Tsao and Steinberg (1988) have very recently reported extending the technique to process coherent radiation fields, thereby reducing speckle and sidelobe artefacts in microwave imagery.

In contrast to the inverse, Wiener and homomorphic filters, which can be considered to be multiplicative filters operating entirely in Fourier-space, CLEAN is a subtractive procedure operating in image-space. In Högbom's terminology $g(\mathbf{x})$, $\hat{h}(\mathbf{x})$ and $\hat{g}(\mathbf{x})$ are called the *dirty map*, the *dirty beam* and the *clean map* respectively. Thus, the CLEAN deconvolution process can be expressed as (Fright 1984, §2.2.3):

> "CLEAN the dirty map (image) with the dirty beam to produce the clean map."

CLEAN is an iterative procedure that can be described, at the i^{th} iteration, by the following steps

$$\hat{f}_{i+1}(\mathbf{x}) = \hat{f}_i(\mathbf{x}) + \beta g_i(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_m), \tag{3.13}$$

and

$$\hat{g}_{i+1}(\mathbf{x}) = \hat{g}_i(\mathbf{x}) - \beta\hat{h}(\mathbf{x} - \mathbf{x}_m)\frac{g_i(\mathbf{x}_m)}{\hat{h}(0)}, \tag{3.14}$$

where β is a real constant called the *loop gain* and $\mathbf{x}_m$ is the coordinate of the pixel of $g_i(\mathbf{x})$ with the largest magnitude, i.e.

$$|g_i(\mathbf{x}_m)| = \max_{\mathbf{x}} g_i(\mathbf{x}). \tag{3.15}$$

Additionally, the initial conditions $f(\mathbf{x}) = 0$ and $g_0(\mathbf{x}) = g(\mathbf{x})$ must be specified.

Iterations cease when the *residual* dirty map $g_i(\mathbf{x})$ is reduced down to what is estimated to be the contamination level, at which point the algorithm is considered to have converged. Irrespective of the contamination level, a necessary, although not sufficient, condition for convergence is that

$$0 < \beta < 2, \tag{3.16}$$

and it is usual for β to be chosen to be less than unity. The final stage of the CLEAN algorithm involves convolving the clean map with a clean beam to enhance the appearance of the reconstruction. The clean beam typically has the same effective width as, but lower sidelobes than, the dirty beam.

Subtractive deconvolution is particularly appropriate when $f(\mathbf{x})$ is composed of a discrete number of isolated, unresolvable parts (Roddier 1988, §5.2.2) and when the main effect of $h(\mathbf{x})$ is to reduce the contrast rather than the resolution (Bates and McDonnell 1986, §17). An example of such a situation is a radio-astronomical telescope observing regions of the heavens containing isolated bright radio sources of small angular extent. CLEAN is also capable of deconvolving blurred images when the Fourier-space samples do not lie on a regular grid (Fright 1984, p. 35). The performance of the CLEAN algorithm is compared with that of the Wiener filter in the example presented in Fig. 3.5.

3.2 Homomorphic Ensemble Blind Deconvolution

Following the definition introduced in §1.5 the ensemble blind deconvolution problem is posed as:

> "Given the ensemble $g_m(\mathbf{x}) = f(\mathbf{x}) \odot h_m(\mathbf{x}) + c_m(\mathbf{x})$; $m = 1, \ldots, M$, of differently blurred versions of the object, retrieve $\hat{f}(\mathbf{x})$, an estimate of $f(\mathbf{x})$."

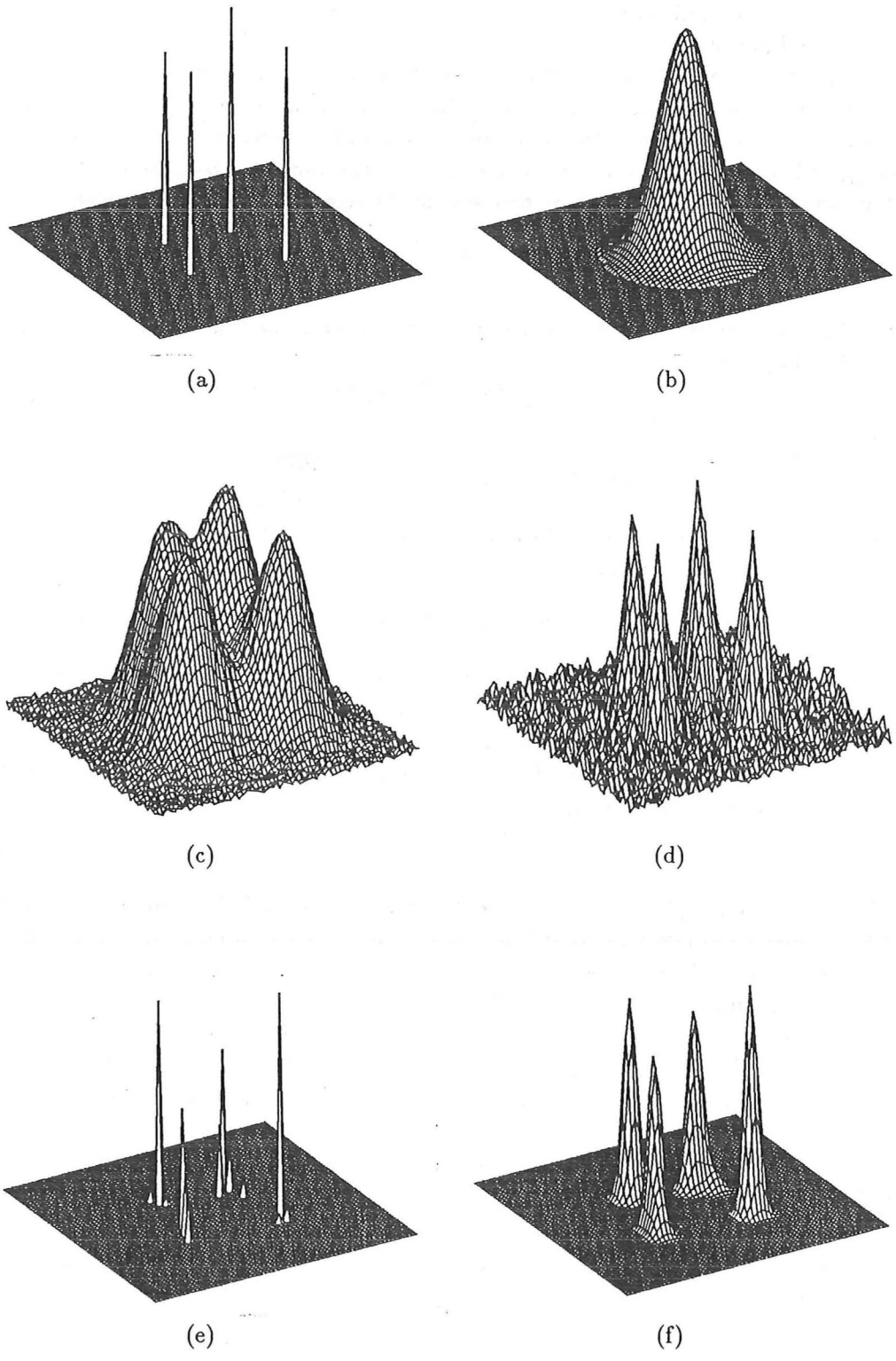


Figure 3.5: A comparison of Wiener filtering and CLEANing of a contaminated convolution of an object comprising four unresolved points: (a) $f(\mathbf{x})$; (b) $h(\mathbf{x})$; (c) the dirty map $g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x}) + c(\mathbf{x})$ ($\epsilon_{\rm dB} = -30$dB); (d) the result of Wiener filtering $g(\mathbf{x})$ with $h(\mathbf{x})$; (e) the clean map after 20 iterations with $\beta = 0.5$ and using $h(\mathbf{x})$ as the dirty beam; (f) the final clean image formed by convolving the clean map (e) with a gaussian clean beam. The structure of $f(\mathbf{x})$ is clearly apparent in both (d) and (f).

In the early 1970s two algorithms for implementing ensemble blind deconvolution were reported in quite separate fields (Bates 1982b, §14). One of these was high angular resolution astronomy, for which Labeyrie (1970) realised that ensemble blind deconvolution techniques, applied to a sequence of short exposure images of a star, could effectively overcome the severe degradation caused by the earth's atmosphere. His, and the most popular of the techniques that have subsequently been developed in this field, are briefly reviewed in Chapter 4, and a new technique is introduced in Chapter 6. The other field, signal processing, is discussed in this section.

A particular application of ensemble blind deconvolution, which was explored by Stockham, Jr. (1971) (*cf.* Stockham, Jr. *et al.* 1975) at roughly the same time as Labeyrie came up with the idea of speckle interferometry, involved restoring early gramophone recordings that had been made on wax disks. The most critical aspect of such a restoration is the removal of artefacts induced by resonances of the primitive recording equipment. These resonances can be usefully modelled as a linear transfer function, which is convolved with the signal to be recorded — in Stockham's situation, the singer's voice.

Had the transfer function of the actual equipment that made the original recording been available, a conventional deconvolution problem would be posed. However, the psf characterizing the device was never known (because the concept was as yet unenvisaged) and has anyway long been lost. Thus, both the psf and the signal are unknown and must be recovered. The restoration technique proposed by Stockham involves forming an estimate of the psf and then applying a conventional deconvolution technique (§3.1) to remove the effect of the psf, thereby restoring the signal. Thus, since the first stage of the algorithm involves estimating the psf, one can consider that the roles of the signal and the psf are reversed.

Stockham, Jr. *et al.* (1975) note that if several recordings of significantly different singers singing different songs, but all made with the same equipment, were available, the psf could be estimated by averaging the cepstra of the recordings (3.9). According to the central limit theorem, the average should converge to $\ln[H(\mathbf{u})]$. Unfortunately, such an ensemble is not available and, anyway, the transfer function differed from recording to recording due to tuning of the equipment by recording engineers.

Alternatively, if the singing can be considered to be nonstationary, it is possible to partition $g(\mathbf{x})$ into an ensemble $g_m(\mathbf{x})$; $m = 1, \ldots, M$ of different musical passages, here denoted by $f_m(\mathbf{x})$, that are each distorted by the same $h(\mathbf{x})$. It is implicit in this partitioning that $h(\mathbf{x})$ is of considerably smaller temporal extent than each of the $g_m(\mathbf{x})$. Upon introducing the notation

$$\langle\cdot\rangle_M = \frac{1}{M}\sum_{m=1}^{M}(\cdot) \tag{3.17}$$

to denote the ensemble average, it follows that

$$\langle \ln[G_i(\mathbf{u})]\rangle_M \approx \langle \ln[F_i(\mathbf{u})]\rangle_M + \ln[H(\mathbf{u})], \tag{3.18}$$

since $h(\mathbf{x})$ is not dependent upon m. If the first term on the RHS converged to zero for large M, one would be left with $\ln[H(\mathbf{u})]$. Unfortunately, this term does not converge to zero. It turns out, however, that a modern recording of a singer singing the same song allows this term to be estimated because the more sophisticated recording equipment, which is now standard, introduces virtually no distortion. Thus, identical processing applied to a modern recording provides an estimate of $\ln[H(\mathbf{u})]$. This can be subtracted from $\ln[G(\mathbf{u})]$ to leave an estimate of $\ln[F(\mathbf{u})]$ from which a reconstruction of $f(\mathbf{x})$ can be formed (*cf.* (3.10) and (3.11). Stockham, Jr. *et al.* (1975)

report that, when they apply their technique to the early recordings of the legendary tenor Enrico Caruso, the clarity of expression is improved and the distortions due to resonances in the recording equipment are significantly reduced.

Stockham's technique makes no attempt to estimate the true phase of the signal, since it is unimportant for audio restorations, because the ear is not sensitive to it (Wang and Lim 1982). The phase is critical for many other applications (Oppenheim and Lim 1981) however. For example, it is shown in Chapter 5 that phase estimation is essential for many image processing applications. In a restricted class of situations for which the psf is symmetric, the phase of the spectrum is directly deducible from its magnitude. Under such conditions Stockham, Jr. *et al.* (1975) report applying their processing to two-dimensional images degraded with limited classes of blurs, namely motion blur and defocus (§1.4).

It is interesting to note that the term blind deconvolution, as invoked in this thesis (see §1.5), is considerably more general than envisioned by Stockham, Jr. *et al.* (1975). Their technique is not strictly blind according to this thesis's definition because they require an ensemble of what effectively are differently distorted recorded signals plus a modern recording of the same musical work to act as a 'reference'. In comparison, the methods introduced in Chapter 7 can successfully remove the distortion from a single recorded signal without requiring any additional *a priori* information.

3.3 Inherent Limitations of Deconvolution

In many practical situations, the blurring and recording of the image cause information about the object to be irretrievably lost. Because of this one cannot hope to recover the components of the convolution without additional *a priori* information. Thus, this loss of information presents a fundamental bound on how faithful reconstructions from a deconvolution algorithm operating on the blurred image can possibly be. Two such limitations are identified and are discussed here.

The first limitation arises when a part of the information about one of the components is missing in the convolution. This may be due, for example, to a pixel of the spectrum of one of the components of the convolution being essentially zero. It follows that, since the spectrum of the convolution is the product of the two components' spectra (*cf.* Table 2.3, convolution theorem), the corresponding pixel in the convolution's spectrum is also zero. Consider the particular point in Fourier-space identified by the position vector $\mathbf{u}'$, which is such that $H(\mathbf{u}') = 0$. It follows that

$$\begin{aligned} G(\mathbf{u}') &= F(\mathbf{u}')H(\mathbf{u}') \\ &= 0, \end{aligned} \tag{3.19}$$

irrespective of the value of $F(\mathbf{u}')$. Thus, no information about the spatial frequency content of $F(\mathbf{u})$ corresponding to the frequency $\mathbf{u}'$ is contained in $G(\mathbf{u})$. Now consider what transpires when one attempts to derive $F(\mathbf{u})$ from $G(\mathbf{u})$ by inverse filtering (see §3.1.1). Since $H(\mathbf{u})$ appears in the denominator of the inverse filter $\Omega(\mathbf{u})$, defined by (3.3), and since $H(\mathbf{u}') = 0$, it follows that $\Omega(\mathbf{u}')$ is undefined. Consequently no estimate of the content of $f(\mathbf{x})$ for the spatial frequency $\mathbf{u}'$ can be made by invoking this filter. In fact, no reconstruction is possible unless an *ad hoc* rule is introduced for endowing the filtered spectrum with a finite value at $\mathbf{u}'$. By comparison, the presence of the additional term Φ in the denominator of the Wiener filter (§3.1.2) permits $W(\mathbf{u})$ to be defined at $\mathbf{u}'$. Thus, it is possible to generate a reconstruction by invoking (3.7) and (3.8). However, it does not follow that the reconstruction is necessarily faithful since $G(\mathbf{u}')$ contains no information about $F(\mathbf{u}')$. It is perhaps

appropriate to regard the Wiener filter as a systematic replacement for the *ad hoc* rule mentioned above.

A second limitation arises because of the unavoidable presence of contamination in real-world images. As discussed in §2.1.3, it is often useful to consider the contamination to have the same probability distribution at all frequencies, i.e. it is 'white'. Consequently, one would expect the parts of an image's spectrum having smallest magnitude to be most susceptible to contamination. In most real-world images the magnitudes of pixels corresponding to low spatial frequencies tend to be considerable larger than those corresponding to high ones. This is apparent in Fig. 3.1(b) and (d) where the low frequencies, here corresponding to the centre of the spectrum, are obviously much brighter than the high frequencies, which occupy the extremities of the spectrum.

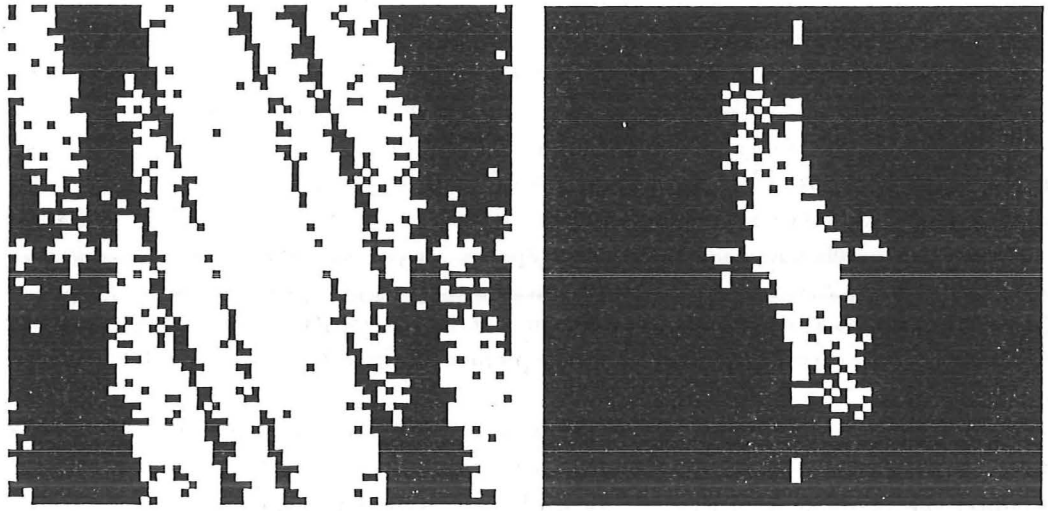
The convolution of images corresponds to the multiplication of their spectra, as emphasized by (3.2). It is informative to investigate the effect this multiplication has upon the magnitude of the spectrum of the convolution, and therefore upon the relative dominance of contamination at different frequencies. The magnitudes of the spectra of typical images all tend to be larger at lower than at higher spatial frequencies. Thus, when two spectra are multiplied together, the magnitudes of the corresponding pixels tend to be similar on the average, so the multiplication is approximately a squaring operation. As regards the spectrum of the convolution, the pixels with small values become even smaller relative to those with large values. The suppression of the high spatial frequencies corresponds to a smoothing of the convolution. Inspection of Fig. 3.1 reveals that the convolution (Fig. 3.1(e)) is smoother than either of the component images (Fig. 3.1(a) and (c)). The reduction in the magnitude of the high spatial frequency components means that the relative effect of contamination is aggravated at higher spatial frequencies, thus leading to a more severe degradation of these frequencies in the convolution than in the component images.

When discussing the concept of domination of a spectrum by contamination, I have found it useful to introduce the concept of the *contamination mask* of a spectrum. The contamination mask for an arbitrary spectrum $Q(\mathbf{u})$ is defined by

$$\Xi[Q(\mathbf{u})] = \begin{cases} 1 & \text{if } E[|C(\mathbf{u})|] < |Q(\mathbf{u})| \\ 0 & \text{otherwise,} \end{cases} \tag{3.20}$$

where $E[\cdot]$ signifies expected (average) value. Thus, $\Xi[Q(\mathbf{u})]$ is a binary mask in the sense that its pixel values are either 0 or 1. Consider the spectrum of the uncontaminated convolution displayed in Fig. 3.1(f). The contamination masks for this spectrum for two levels of contamination are depicted in Fig. 3.6. The regions of the mask set to unity reveal the parts of Fourier-space that are most likely to contain useful data, whereas those areas corresponding to the mask being zero are considered to contain no useful information.

Before deconvolution algorithms can be applied with confidence, their efficacy when processing degraded data must be evaluated. When performing such an evaluation it is useful to recognize and incorporate fundamental limitations which would apply to all deconvolution algorithms for a given set of degraded data. This is particularly meaningful when comparing algorithms with the aid of computer simulations, since one can then accurately control the level and nature of the contamination added to the convolution. In a computer simulation, one knows the convolution exactly, so it is possible to accurately compute the contamination mask (3.20) of the convolution's spectrum. This mask reveals the regions of Fourier-space in which the true



(a) (b)

Figure 3.6: Contamination masks (see (3.20)) of the convolution $g(\mathbf{x})$ shown in Fig. 3.1(e) for: (a) $\epsilon_{\text{dB}} = -40\text{dB}$; (b) $\epsilon_{\text{dB}} = -20\text{dB}$. It is apparent that regions of Fourier-space corresponding to high spatial frequencies are usually more susceptible to contamination since pixels in those regions have smaller magnitude than those in regions corresponding to low frequencies.

spectrum of the convolution is expected to dominate the contamination, and thus contain useful information.

Deconvolution algorithms can not be expected to recover information corresponding to frequencies that have been effectively lost in the convolution. Consequently, the contamination mask for the convolution is also an effective mask for each of the components. When investigating the effectiveness of different deconvolution algorithms, it is useful to take as benchmarks (representing the best possible reconstructions that an 'ideal' algorithm could generate) degraded versions of the components, formed by windowing each component's spectrum with the convolution's contamination mask. Therefore, the fidelity of any reconstructed image is appropriately assessed by comparing it with its 'band-limited' version. This approach is invoked for evaluating the deconvolution algorithms presented in Chapter 7.

Chapter 4

Astronomical Speckle Imaging

Often, when an image needs deblurring, detailed information about what caused the blurring is not available, so the conventional deconvolution techniques described in §3.1 cannot be applied. To deblur the image in such a situation one must resort to blind deconvolution techniques. The term *blind deconvolution* means, as explained in §1.5, deconvolution when the psf is not known *a priori*.

The techniques described in this chapter have been developed for overcoming the 'seeing problem' (§4.1), i.e. for combating blurring that degrades high resolution astronomical images captured by earth-based telescopes. It is shown in §4.2 that these techniques implement *ensemble blind deconvolution*, where, following the definition given in §1.5, the ensemble blind deconvolution problem is posed as:

> "Given only $g_m(\mathbf{x}) = f(\mathbf{x}) \odot h_m(\mathbf{x}) + c_m(\mathbf{x})$, for $m = 1, \ldots, M$, recover $\hat{f}(\mathbf{x})$, an estimate of $f(\mathbf{x})$."

The more general blind deconvolution problem, which arises when only a single blurred image is available, is discussed in Chapter 7.

In 1970, Labeyrie reported a new technique for overcoming the distortion of astronomical images caused by turbulence in the earth's atmosphere (Labeyrie 1970). This turbulence gives rise to random fluctuations in the atmosphere's refractive index. Labeyrie realized that a short duration exposure effectively freezes these fluctuations. The resulting image, although blurred, retains information out to the diffraction limit of the telescope. By appropriately processing many short exposure images, each of which is differently blurred, high angular resolution information about the object is recoverable. Following Labeyrie's initial proposal of speckle interferometry slightly less than 20 years ago, several groups began researching this field, which is now known as *astronomical speckle imaging* (Bates 1982b). Many successful algorithms have resulted, the most popular of them being reviewed in this chapter. Some terminology appropriate for discussions of celestial objects is presented in §4.3.

Like speckle interferometry, several of the astronomical speckle imaging algorithms involve averaging quantities calculated from the spectra of the speckle images (speckle spectra). Such algorithms are here said to be Fourier-space techniques. For instance, in speckle interferometry, which is described in §4.5, the quantities that are averaged are the squared magnitudes of the speckle spectra. Unfortunately, speckle interferometry discards the Fourier-space phase, so only the autocorrelation of the image is directly recoverable. Fourier-space techniques that involve more sophisticated processing to preserve the Fourier-space phase have been devised. The most popular of these, Knox-Thompson (Knox and Thompson 1974) and triple correlation (Lohmann *et al.* 1983) processing, are described in §§4.6 and 4.7.

Other, image-space, astronomical speckle imaging algorithms operate on the

speckle images themselves, rather than their spectra. Shift-and-add (Bates 1976; Bates and Cady 1980) is one technique that is finding increasing application in observatories throughout the world. Shift-and-add is both conceptually and computationally simple. Many examples of blind deconvolution presented in Chapter 7 are related to this technique. Consequently, the discussion of shift-and-add presented in §4.8 and other sections of this thesis goes into considerably more detail than do the accounts of other astronomical speckle imaging techniques.

Zero-and-add (Davey *et al.* 1986; Sinton 1986) is a relatively new astronomical speckle imaging technique. Zero-and-add is essentially a one-dimensional technique although extensions allowing it be applied to two-dimensional data have been devised (Bates and Lane 1987b). Sinton *et al.* (1986) (*cf.* Davey *et al.* 1988a) have indicated how zero-and-add might be applied in conjunction with shift-and-add to improve the quality of the image resulting from shift-and-add. Since Chapter 6 is devoted to describing zero-and-add, the technique is not discussed further in this chapter.

Speckle imaging techniques have been extended to infrared wavelengths (Sibille *et al.* 1979; Howell *et al.* 1981), thereby taking advantage of the inherent resolving power of telescopes that operate at these wavelengths. Thus, regions of the heavens that radiate at infrared wavelengths, and that are often of great interest to astrophysists and astronomers (Dyck *et al.* 1984), can be observed in greater detail. A brief description of the commonly adopted procedure for recording infrared speckle images is presented in §4.9, since it is somewhat different from that used at visible wavelengths. There is further discussion of infrared speckle imaging in §6.3.

For more complete details of astronomical speckle imaging techniques and applications the reader is directed to several comprehensive reviews of the topic (Bates 1982b; Dainty 1984; Sinton 1986; Roddier 1988). Because of the existence of these reviews, only the essence of each of the speckle imaging techniques that are proving most popular is discussed here. For the reader interested in the effects of the atmosphere on optical propagation and astronomy, several review articles have been written, both from theoretical (Strohbehn 1971; Roddier 1981) and observational (Woolf 1982), viewpoints.

Detailed treatment of photon noise has not been included in this chapter. This may seem surprising at first since, as outlined in §4.2, it is the dominant form of contamination in most astronomical speckle imaging contexts. However, there are several reasons justifying this omission:

(i) Various aspects of photon noise, as it affects astronomical speckle imaging, has been recently treated by several authors (see, for example, Goodman and Belsher 1976; Dainty 1978; Dainty 1984, §7.3; Wirnitzer 1985; Northcott *et al.* 1988; Ayers *et al.* 1988).

(ii) It is worthwhile attempting to complete the theory of bright-object image formation as far as one can, because it is of value in itself, and because speckle imaging is useful (and potentially more so perhaps) in fields other than astronomy (for example, ultrasonic imaging (Bates and Minard 1984), synthetic aperture radar (Eichel *et al.* 1989), and speech processing (Brieseman *et al.* 1987)).

(iii) The usage of the term contamination in this thesis implies all discrepancies between the model and reality, although it is possibly less useful for faint objects, since, for astronomical imaging, photon noise is often more severe than other forms of contamination.

4.1 The Seeing Problem

In conventional astronomical imaging, celestial objects are viewed from the earth with the aid of a telescope. The two-dimensional image formed by the telescope is integrated over time using photographic film, or some sort of electronic camera. The length of exposure depends upon the brightness of the object and size of the telescope, but exposure times of several minutes, or even hours, are not uncommon. To make the exposure, the telescope must track the object across the sky gathering radiation either emitted (e.g. by stars), or reflected (e.g. by nebulae or planets), by that object. Fig. 1.1 depicts an idealized representation of the optical astronomical image recording situation.

The angular resolution of a perfectly figured optical telescope is limited by diffraction to (Goodman 1968)

$$\theta = \frac{1.22\lambda}{D}. \tag{4.1}$$

θ, an angle expressed in radians, is the smallest angular separation between features of the object that the telescope is able to detect. D and λ are both conveniently expressed in metres (m), and are the diameter of the telescope and the wavelength of radiation incident upon the telescope respectively. Thus, for $\lambda = 5 \times 10^{-7}$m (i.e. $\lambda = 500$nm or green light), a 4m telescope resolves to about 0.025arcsec.

In practice, however, inhomogeneities in the refractive index of the earth's atmosphere severely degrade the resolution of large earth-based telescopes (see Fig. 4.1). Such inhomogeneities are due to changes in humidity, the winds and thermal effects (Roddier 1981, §2). They cause the planar wavefront incident upon the atmosphere to be distorted during its passage through the atmosphere. The magnitude of the wavefront is not greatly distorted although its slight fluctuations, called *scintillation* (Roddier 1981, §7), generate the twinkling seen when observing stars. The wavefront is no longer planar, but becomes unpredictably curved, so that different parts of it are differently delayed by the time the wavefront reaches the telescope pupil. The result is that the pupil field exhibits appreciable phase variations over distances that tend to be much smaller than the diameter of a large telescope. These seemingly random phase variations cause the incoming radiation to interfere at the focal plane of the telescope. The interference tends to generate many versions of the object, each having a random amplitude and position, that move unpredictably with time (Bates 1982b). The long exposure image is simply the time average of these interference patterns.

The atmosphere can be characterized, at a given frequency, by the *Fried seeing parameter* r_0 (Fried 1966; Roddier 1981, §4.3). It is useful to consider r_0 to be the average diameter of 'blobs' (often termed *seeing cells*) in the turbulent medium. All parts of a particular seeing cell can be considered to introduce approximately the same phase delay to radiation passing through it. Like other properties of atmospheric turbulence, r_0 varies from place to place and time to time, although typical values at visible wavelengths are of the order of 10cm, with variations up to an order of magnitude on either side of this (*cf.* Labeyrie 1976; Woolf 1982, §4; von der Lühe 1984). r_0 can change significantly over a time scale of hours or even minutes (Mariotti 1983).

All the radiation entering the pupil of a small telescope (i.e. $D \leq r_0$) can be considered to have passed through the same seeing cell, and therefore, to have been delayed similarly. Consequently, the image formed by such a telescope suffers negligible loss of resolution, although it does tend to wander in the focal plane. The pupil of a large telescope can be partitioned into many sub-pupils, each of diameter r_0,

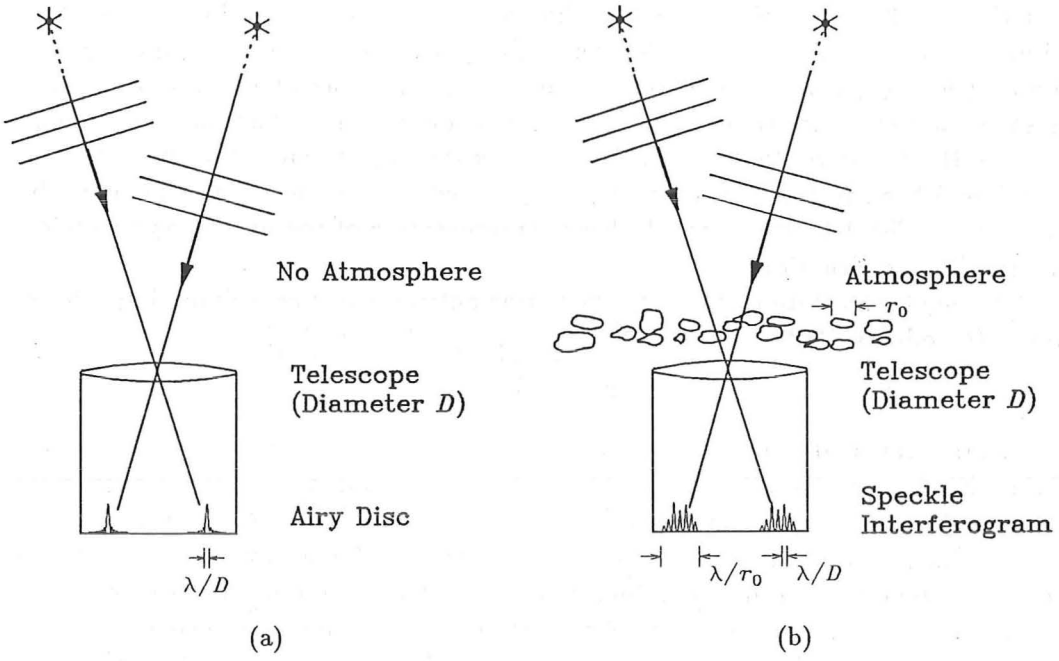


Figure 4.1: The degrading influence of turbulence in the earth's atmosphere: (a) the image recorded by a telescope possessing no aberrations in the absence of turbulence would be an Airy disc of size λ/D; (b) in reality, speckle patterns, which are formed in the focal plane of a large telescope, are much larger than the Airy disc. However, the typical speckle size is approximately the same as that of the Airy disc.

within each of which there is negligible distortion. However, there is a markedly different phase delay between the radiation collected from the different sub-pupils. The image formed by such a telescope under such conditions exhibits severe degradation.

The long-exposure image of a stellar object is called the *seeing disc*. The diameter of the seeing disc is dependent upon the properties of the atmosphere (i.e. upon the *seeing*), but a typical diameter at a good observation site in the visible range is of the order of an arc-second (Dainty 1984, §7.2.2; Roddier 1988, §1.1). This is equivalent to the resolving power of a telescope having a diameter of approximately r_0. Thus, a 4m telescope operating under these conditions, would have a resolving power approximately 40 times less than its theoretical diffraction limit. This degradation in the resolution of stellar images, caused by the earth's atmosphere, is called the *astronomical seeing problem* (Bates 1982b, §2). Even with the resolution limit imposed by the atmosphere, large telescopes do still have the advantage of collecting more light, allowing fainter objects to be observed.

4.2 Astronomical Speckle

If the duration of an exposure is very short, the motion of the atmosphere is effectively frozen and the resulting image has a speckly appearance (Labeyrie 1970; Bates 1982b) reminiscent of laser speckle (Goodman 1976; Dainty 1984). Such an image is consequently called a *speckle image* (or specklegram or speckle frame). A typical speckle image of a bright star, which is unresolvable (§1.1) by the telescope,

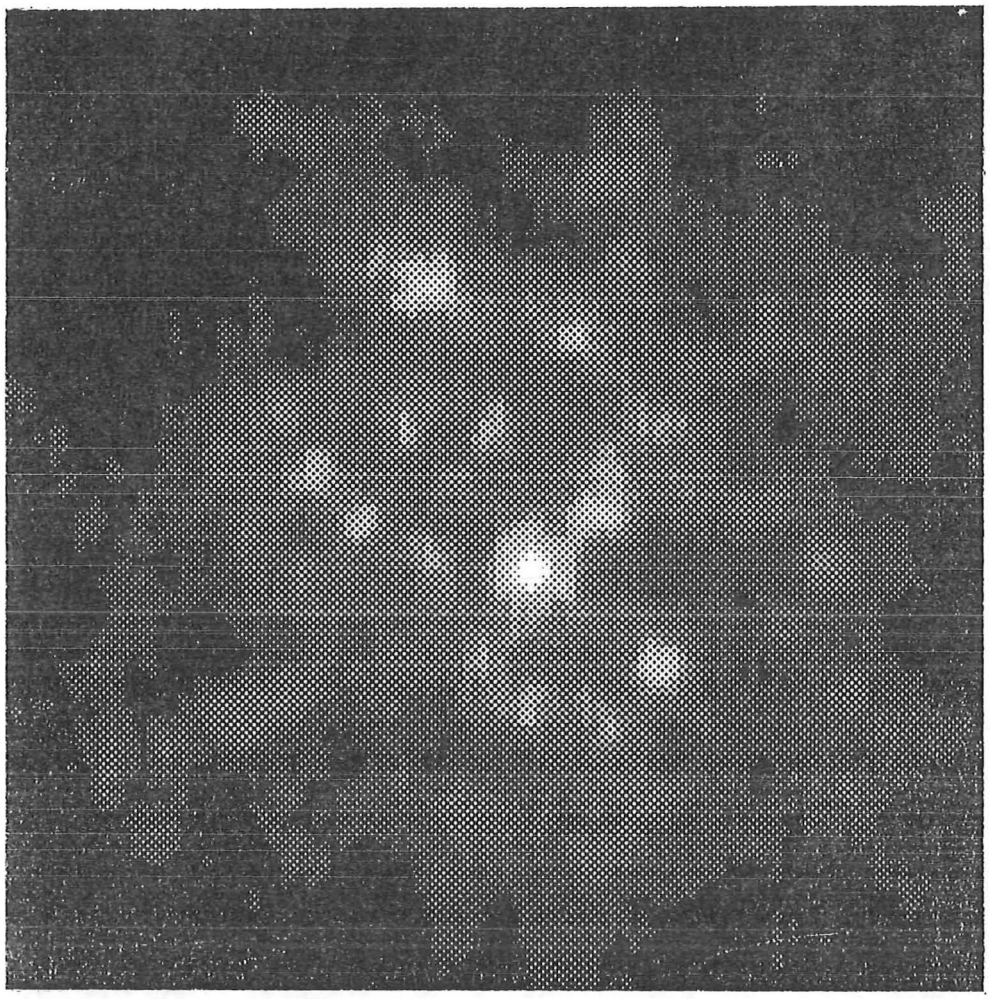


Figure 4.2: A typical short exposure speckle image of the star Bellatrix (magnitude 1.64), otherwise known as γ Orionis. Together with Betelgeuse, this star marks the shoulder blades of the great mythological celestial warrior, Orion. ($D = 4$m, $\lambda_0 = 550$nm and $\Delta\lambda = 20$nm.) Image courtesy of Steward Observatory.

is displayed in Fig. 4.2. Provided certain conditions (Roddier 1988, §2.3), which are discussed in following paragraphs, are satisfied, a speckle image can be approximately, but usefully, expressed as a contaminated convolution (Bates 1982b, §3.2). The components of the convolution are the object and a psf characterizing the atmosphere above the pupil during the recording of the speckle image. This psf typically resembles a collection of randomly positioned and scaled Airy discs (§1.1), shown in idealized form in Fig. 1.6(c). Since an Airy disc is (as far as imaging with a particular telescope is concerned) equivalent to a two-dimensional impulse, and since the convolution of a function with an impulse only results in a shifting and scaling of the function (see §2.2), a speckle image can usefully be considered to be a collection of randomly positioned and scaled versions of the (diffraction-limited) object.

The first condition which must be met for the convolution relation to be valid is that the radiation being collected by the telescope must be filtered to a sufficiently narrow bandwidth before being recorded. Roddier (1988, §2.3.2) gives the following

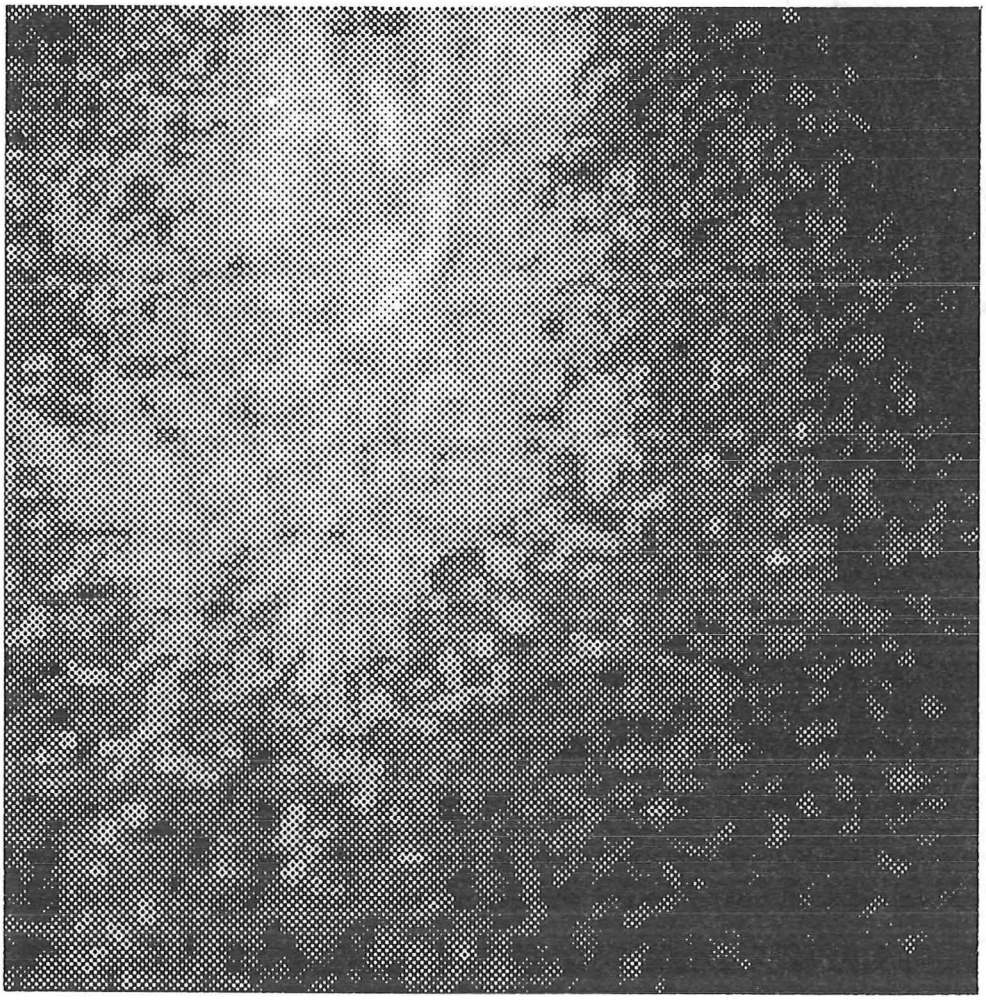


Figure 4.3: A wide bandwidth speckle image of the binary star ADS 2200. Note that although a speckly nature is apparent near the image centre, the wide bandwidth used to produce this image results in the observable radial smearing of those speckles which occur distant from the image centre. ($D = 1.88$m, $\lambda_0 \approx 450$nm, $\Delta\lambda = 300$nm, $\tau_s = 8$ms.) Image courtesy of University of Hokkaido.

rule of thumb for the bandwidth

$$\frac{\Delta\lambda}{\lambda_0} \ll \frac{r_0}{D}, \tag{4.2}$$

where $\Delta\lambda$ is the wavelength bandwidth and λ_0 is the wavelength at the centre of the band. One form of degradation of speckle image quality, which results if the images are recorded with too large a bandwidth, is radial smearing (Bates and Cady 1980; Dainty 1984, §7.2.6) such as that apparent in Fig. 4.3.

Secondly, the short exposure duration, τ_s, must be sufficiently short that the turbulence is effectively frozen. If the duration of the exposure is excessively long, the speckles become blurred and the contrast drops (Roddier 1988, §2.3.1). Conversely, since the amount of light collected to make the exposure is proportional to the exposure time, and since in many astronomical imaging situations the object being observed is very faint, τ_s must be made as long as possible. In practice, a compromise that yields an optimum SNR (Dainty 1984, §7.3.2) is usually reached. A typical

Parameter	Scaling factor	500nm (visible)	5μm (infrared)
Fried seeing parameter (r_0)	$\propto \lambda^{6/5}$	15cm	200cm
Bandwidth $\Delta\lambda$	$\propto \lambda^2$	10nm	1000nm
Short Exposure Duration τ_s	$\propto \lambda^{6/5}$	0.02s	0.3s
Isoplanatic Angle	$\propto \lambda^{6/5}$	3–6arcsec	40–80arcsec

Table 4.1: Typical parameters characterizing the atmosphere at visible and infrared wavelengths and their dependency upon wavelength.

value of τ_s at visible wavelengths is 10ms although it can vary significantly either way (Roddier 1988, §2.3.1).

Thirdly, all parts of the object being viewed must lie within the same isoplanatic patch (see §2.2) (Roddier 1988, §2.3.4). Typical values of the isoplanatic patch diameter range from 1.5–5arcsec (Roddier 1988, §2.3.4), although Ebersberger and Weigelt (1985) present measurements which suggest that there can be significant isoplanatism out to approximately 8arcsec.

Finally, for speckles to form, the telescope must be of sufficiently large diameter, typically greater than 0.5m for visible wavelengths. The typical size of an individual speckle in a speckle image is approximately the size of the telescope's Airy disc (Roddier 1981, p. 312). Thus, increasing the size of the telescope decreases the typical speckle size. The average number, N_{sp}, of speckles per image is approximately given by (Bates 1982b, p. 225)

$$N_{sp} \approx \frac{D^2}{r_0^2}, \tag{4.3}$$

since this is the number of seeing cells directly above the telescope. Typical values for the parameters governing the capture of speckle images at both visible and infrared wavelengths are summarized in Table 4.1.

Because the atmosphere is in a state of continual turbulence, a temporal sequence of exposures of a celestial object generates an ensemble of images, each of which is blurred differently. The *redistribution time* τ_r of the atmosphere specifies how long an interval there must be between members of the sequence in order that each is essentially statistically independent (Bates 1982b, §2.4). The m^{th} speckle image in an ensemble of M such images is written

$$s_m(\mathbf{x}) = f(\mathbf{x}) \odot h_m(\mathbf{x}) + c_m(\mathbf{x}), \tag{4.4}$$

where $h_m(\mathbf{x})$ is the psf characterizing the atmosphere during the m^{th} exposure. A long exposure image can be thought of as the integration of speckle images over many redistribution times, during which the speckles smear out into the featureless seeing disc. Thus, although each speckle image contains frequency information out to the diffraction limit of the telescope λ/D, the seeing disc only has information out to λ/r_0.

The astronomer is eager to capture as much of the radiation as possible to maximize the signal-to-noise ratio (SNR) of the imaging process. Since almost all objects of interest to astronomers are extremely distant from earth, the level of light collected by the telescope tends to be very low. This problem is largely overcome in conventional astronomy by integrating the light for a lengthy period. However, when implementing speckle imaging techniques, this problem is aggravated by the constraints of narrow bandwidth and short duration exposures. Consequently, when imaging faint

objects, photon noise is the dominant type of contamination (Dainty 1984, §7.3; Roddier 1988, §2.5). However, sources of contamination which are common to other forms of image processing also occur when recording astronomical images. These include noise induced by the measurement apparatus, and thermal noise. Dainty (1974) and Roddier (1988, §2.5) discuss the SNR and limiting magnitude of astronomical speckle imaging. Short reviews of practical implementations of speckle recording apparatus are also provided by Dainty (1984, §7.5) and Roddier (1988, §2.6).

Since the goal of any astronomical speckle imaging technique is to recover an estimate of the object from the ensemble of speckle images, it follows that these techniques implement ensemble blind deconvolution (§1.5).

4.3 Terminology for Celestial Objects

It is useful here to introduce terminology for describing classes of astronomical object that are mentioned in this chapter. Reference should also be made to the glossary of general astronomical terminology presented prior to Chapter 1.

An *unresolved object* (*cf.* §§1.1 and 4.2) is one that has no detail detectable by the particular telescope which is making the observation. Such an object is, therefore, essentially an impulse (§2.2).

A *multiple star* is a collection of several *component* stars, each of which are here assumed to be unresolvable (note that all stars, apart from a few giant stars in our galaxy, are unresolvable with even the largest existing telescopes). The separations of the component stars are assumed to be resolvable. Thus, a general multiple star comprising J unresolvable components can be described by a sum of weighted impulses i.e.

$$m_J(\mathbf{x}) = \delta(\mathbf{x}) + \sum_{j=1}^{J-1} R_j \delta(\mathbf{x} - \mathbf{s}_j), \tag{4.5}$$

where $m_J(\mathbf{x})$ denotes the star, and R_j and

$$\mathbf{s}_j = |\rho_j| \mathrm{e}^{\imath\theta_j} \tag{4.6}$$

are here called the brightness ratio and vector separation of the j^{th} component. ρ_j and θ_j are termed the separation and position angle of that component. The first term on the RHS of (4.5) describes what is called the *primary* component, and is usually the brightest component. Thus, because the brightness ratios must be positive, since the radiation is incoherent (§4.1), it follows that all the brightness ratios are in the range 0 to 1. Combining (4.5) with (4.4) reveals that the m^{th} speckle image of a multiple star can be written as

$$s_m(\mathbf{x}) = h_m(\mathbf{x}) + \sum_{j=1}^{J-1} R_j h_m(\mathbf{x} - \mathbf{s}_j). \tag{4.7}$$

A *binary star* (or simply a binary) is a multiple star having two components. When describing a binary, the summation in the RHS of (4.5) involves only a single term, so it is convenient to drop the subscript j. Thus, for example, the position angle of a binary is written as θ and the vector separation as $\mathbf{s}$.

An *extended object* is one that exhibits structure detectable by the telescope making the observation. Examples of extended objects are a few supergiant stars close to earth (notably Betelgeuse), nebulae, asteroids, the planets and, of course, our sun.

4.4 Computation of Simulated Speckle Images

Several computer simulations of astronomical speckle imaging are presented in this, and subsequent chapters, of this thesis. The way in which the speckle images were generated for these simulations is briefly described in this section. Since the procedures invoked to generate one-dimensional and two-dimensional speckle images differ slightly, the two cases are treated separately in §§4.4.1 and 4.4.2.

4.4.1 One-Dimensional Speckle Images

The following steps describe the procedure followed to generate a single contaminated one-dimensional speckle image (*cf.* Sinton *et al.* 1986, §4.4):

(i) A one-dimensional array of pixels having real correlated values generated according to the first-order Markov process described by

$$\phi(u_1) = \beta_1 \tag{4.8}$$

$$\phi(u_n) = \gamma\phi(u_{n-1}) + \sqrt{1-\gamma^2}\beta_n; \quad n > 1, \tag{4.9}$$

where the n^{th} pixel, which has the value $\phi(u_n)$, is situated at u_n. Furthermore, β_n is the n^{th} member in an ensemble of pseudo-random numbers uniformly distributed between $-a$ and a, where a is a pre-set positive constant, and γ is a parameter defining the degree of correlation between pixels. When generating one-dimensional speckle images for the examples presented this thesis, a and γ were chosen to be 15 and 0.995.

(ii) A one-dimensional array of complex valued pixels was created by setting the n^{th} pixel value to be $e^{i\phi(u_n)}$.

(iii) The complex array was inverse Fourier transformed, and the value at each pixel was replaced by its squared magnitude to produce an incoherent speckle psf, $h_m(x)$, typified by Fig. 4.4(b).

(iv) A speckle image was generated as the convolution of the object and this speckle psf. A typical speckle image, formed by convolving the object shown in Fig. 4.4(a) with the speckle psf depicted in Fig. 4.4(b), is displayed in Fig. 4.4(c).

(v) Contamination was then added to the speckle image in the manner described in §2.1.3.

An ensemble of M independent speckle images was generated by repeating the above steps M times, each time choosing a different starting seed for the pseudo-random generator employed in step (i).

Sinton (1986, §4.4) reports that the statistics of the speckle images generated in this manner closely match those of actual speckle images.

4.4.2 Two-Dimensional Speckle Images

Because the Markov process in step (i) of the list in §4.4.1 is readily implemented only in one dimension, alternative methods of simulation must be devised for two-dimensional simulations. In this thesis a single contaminated two-dimensional speckle image was generated by invoking the procedure described by the following steps (*cf.* Minard 1985; Sinton *et al.* 1986):

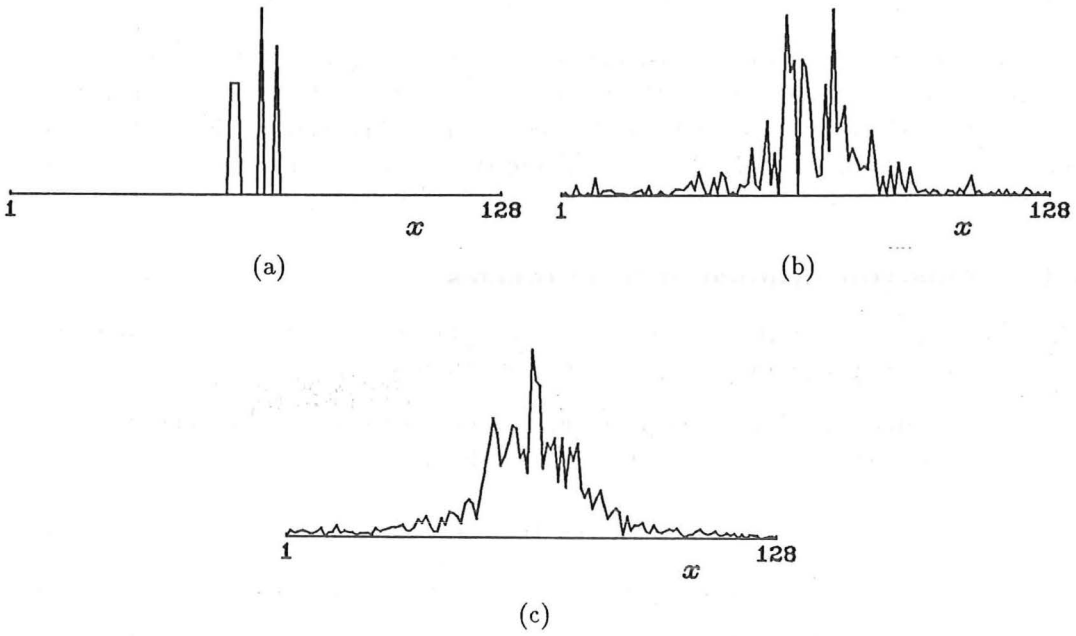


Figure 4.4: Generation of speckle images for computer simulations of one-dimensional speckle imaging: (a) the object $f(x)$; (b) a typical speckle psf $h_m(x)$; (c) the speckle image $s_m(x)$ generated by convolving (a) and (b).

(i) A two-dimensional array of pixels having complex values was generated by setting the magnitude of each pixel to be unity and its phase to be pseudo-random, and uncorrelated with the phases of other pixels, uniformly distributed between $-\pi$ and π.

(ii) The array was inverse Fourier transformed, and the value at each pixel was replaced by its squared magnitude.

(iii) The array was then multiplied by an image characterizing the shape of the seeing disc, to form the speckle psf, typified by Fig. 4.5(b). Following well established theory (Roddier 1981), this image, here denoted by $\mathscr{sd}(\mathbf{x})$, was computed as

$$\mathscr{sd}(\mathbf{x}) = A\mathrm{e}^{-kr^{5/3}}, \tag{4.10}$$

where r is the radial distance of a given pixel from the origin, A is an arbitrary positive constant, and k is a positive constant. The similarity of (4.10) to a Gaussian function (Kreysig 1979, §20.10) is obvious when k is written

$$k = \frac{1}{2\sigma^2}. \tag{4.11}$$

Thus, $\mathscr{sd}(\mathbf{x})$ is similar in form to a Gaussian function, but the 'tails' of $\mathscr{sd}(\mathbf{x})$ do not fall off as rapidly (with increasing r) as those of the corresponding Gaussian function. $\sigma = 8$ was chosen for the simulations presented in this thesis.

(iv) A speckle image was generated by convolving the object and the speckle psf. A typical speckle image, formed by convolving the object shown in Fig. 4.5(a) and the speckle psf depicted in Fig. 4.5(b), is displayed in Fig. 4.5(c).

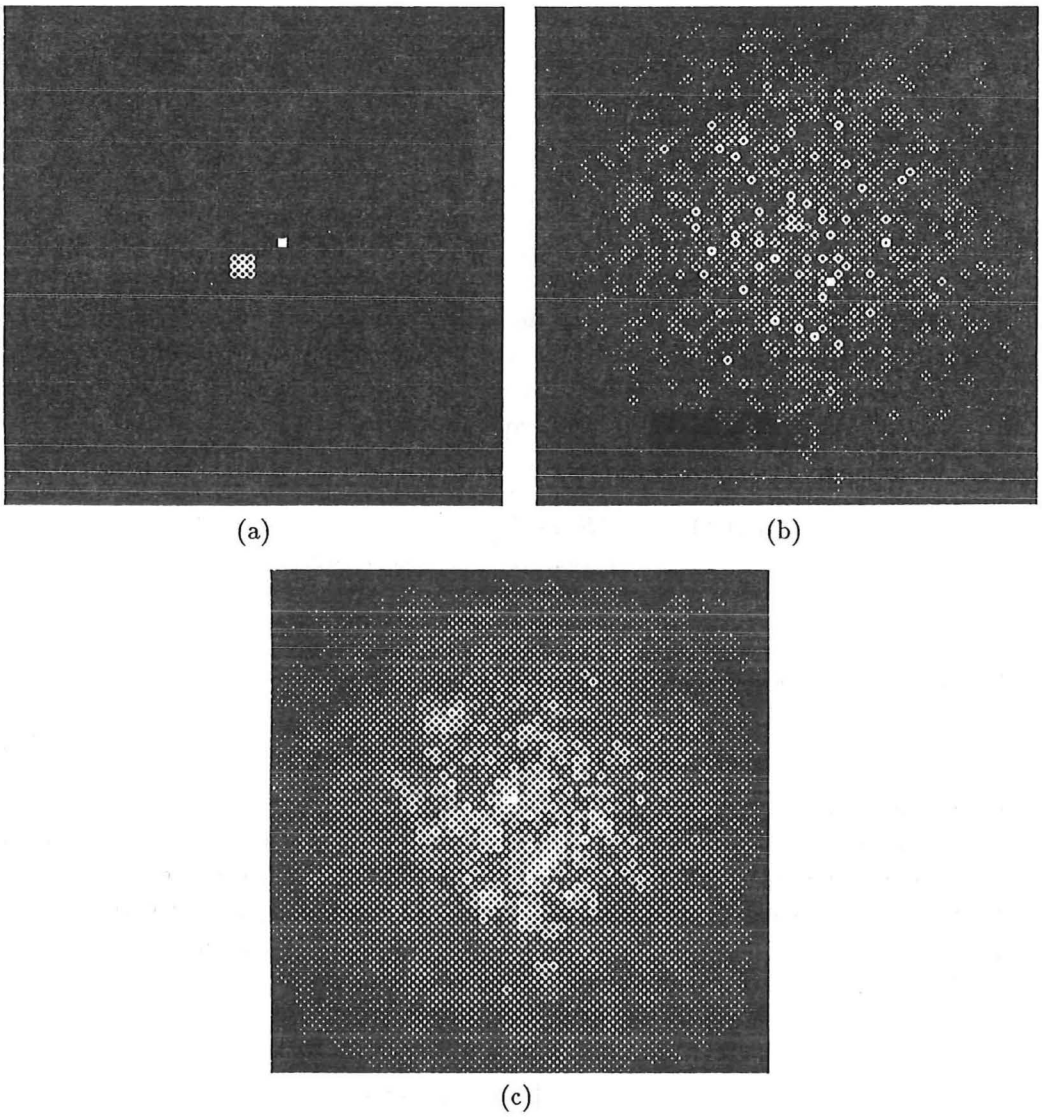


Figure 4.5: Generation of speckle images for computer simulations of two-dimensional speckle imaging: (a) the object $f(x, y)$; (b) a typical speckle psf $h_m(x, y)$; (c) the speckle image $s_m(x, y)$ generated by convolving (a) and (b).

(v) Contamination was then added to the speckle image in the manner described in §2.1.3.

An ensemble of M independent speckle images was generated by repeating the above steps M times, each time choosing a different starting seed for the pseudo-random generator employed in step (i).

4.5 Speckle Interferometry

In 1970, Labeyrie reported the first astronomical speckle imaging technique, which is now known as *speckle interferometry* (Labeyrie 1970). Labeyrie realized that, unlike the seeing disc, a short exposure speckle image contains spatial frequency information out to the diffraction limit of the telescope. It is the process of integrating the speckle images that irretrievably destroys this information in a long exposure image.

However, speckle interferometry provides an alternative technique for averaging the information in each of the speckle images, so that the spatial frequency information content is preserved. Speckle interferometry is now discussed in terms of the notation introduced in (4.4).

Appealing to the convolution theorem (§2.4), the Fourier transform of a speckle image (4.4) is written

$$S_m(\mathbf{u}) = F(\mathbf{u})H_m(\mathbf{u}) + C_m(\mathbf{u}). \tag{4.12}$$

$|S_m(\mathbf{u})|$ typically has significant value out to the diffraction limit of the telescope although the ensemble average (see (3.17)) of the spectra (i.e. $\langle S_m(\mathbf{u})\rangle_M$) is approximately zero for $|\mathbf{u}| > \lambda/r_0$. Labeyrie recognized that, by discarding the phase through computing $\langle |S_m(\mathbf{u})|^2\rangle_M$, spatial frequency information out to the diffraction limit of the telescope is retained. This averaging yields

$$\begin{aligned} S_{\mathrm{LA}}(\mathbf{u}) &= \langle |S_m(\mathbf{u})|^2\rangle_M \\ &= \langle |F(\mathbf{u})H_m(\mathbf{u}) + C_m(\mathbf{u})|^2\rangle_M \\ &= |F(\mathbf{u})|^2 H_{\mathrm{LA}}(\mathbf{u}) + \tilde{C}(\mathbf{u}), \end{aligned} \tag{4.13}$$

where $H_{\mathrm{LA}}(\mathbf{u})$, the *speckle transfer function*, is defined to be

$$H_{\mathrm{LA}}(\mathbf{u}) = \langle |H_m(\mathbf{u})|^2\rangle_M, \tag{4.14}$$

and $\tilde{C}(\mathbf{u})$ incorporates all of the contamination terms.

For given seeing conditions the $H_{\mathrm{LA}}(\mathbf{u})$ is deterministic, and therefore can be measured or predicted (Sinton 1986, §3.3.1). The resulting estimate, here denoted by $\hat{H}_{\mathrm{LA}}(\mathbf{u})$, is then used to inverse filter (§3.1.1), or Wiener filter (§3.1.2), $S_{\mathrm{LA}}(\mathbf{u})$ leaving an estimate of $|F(\mathbf{u})|^2$. When contamination is neglected, this is written

$$\begin{aligned} |\hat{F}(\mathbf{u})|^2 &= \frac{S_{\mathrm{LA}}(\mathbf{u})}{\hat{H}_{\mathrm{LA}}(\mathbf{u})} \\ &= \frac{|F(\mathbf{u})|^2 H_{\mathrm{LA}}(\mathbf{u})}{\hat{H}_{\mathrm{LA}}(\mathbf{u})} \\ &\approx |F(\mathbf{u})|^2. \end{aligned} \tag{4.15}$$

The autocorrelation theorem (§2.4) shows that the inverse Fourier transform of LHS (4.15) yields $\mathcal{A}\left[\hat{f}(\mathbf{x})\right]$.

One standard way of estimating $H_{\mathrm{LA}}(\mathbf{u})$ is to apply speckle interferometry to an ensemble of speckle images of an unresolvable reference object $f_\delta(\mathbf{x}) \approx a\delta(\mathbf{x})$, where a is a constant. There is the requirement, however, that the seeing is statistically similar when these speckle images are recorded to that existing when $f(\mathbf{x})$ was observed. Numerous techniques for compensating for the speckle transfer function without recourse to observing a reference star have been devised (*cf.* Korff *et al.* 1972; Welter and Worden 1978). Given knowledge of the telescope's performance and an estimate of r_0 (von der Lühe 1984), the speckle transfer function can be predicted from theoretical models of optical propagation through the atmosphere (Aime *et al.* 1978).

Appealing to the autocorrelation theorem (§2.4), the image-space version of (4.13) is

$$s_{\mathrm{LA}}(\mathbf{x}) = \mathcal{A}[f(\mathbf{x})] \odot \langle \mathcal{A}[h_m(\mathbf{x})]\rangle_M + \tilde{c}(\mathbf{x}), \tag{4.16}$$

showing that speckle interferometry is equivalent to averaging the image-domain autocorrelation of the speckle images, resulting in a diffraction limited estimate of the object's autocorrelation being reconstructed. Thus, speckle interferometry allows

objects to be resolved to the diffraction limit of the telescope. However, because only the autocorrelation of the object is recovered, speckle interferometry is not explicitly an imaging technique.

Speckle interferometry becomes an imaging technique when it is combined with algorithms that allow the Fourier phase to be recovered from the Fourier magnitude. This recovery of $F(\mathbf{u})$ when $\mathcal{P}[F(\mathbf{u})]$ is unknown, or is only known approximately, is called Fourier phase retrieval (Fright 1984; Bates and Mnyama 1986) and is reviewed in Chapter 5. Thus, by invoking a Fourier phase retrieval algorithm with $|\hat{F}(\mathbf{u})|$ as the input, one is able to reconstruct $\hat{f}(\mathbf{x})$. Other speckle imaging techniques, which explicitly attempt to preserve the phase information present in the speckle images, are described in the following sections.

Speckle interferometry has been employed by observational astronomers for more than fifteen years since its first application by Gezari *et al.* (1972). Many classes of object have been observed, including binary star systems (McAlister 1985; McAlister and Hartkopf 1988), planets (Hege *et al.* 1982) and asteroids (Drummond *et al.* 1985). Difficulties in accurately compensating for the speckle transfer function make accurate photometric information retrieval difficult (Dainty 1984, §7.6). For example, the limb darkening of extended objects and the relative brightnesses of the components of a binary star system are more difficult to obtain than positional information (Sinton 1986).

4.6 Phase Gradient Averaging: Knox-Thompson

Phase gradient averaging algorithms are designed to explicitly preserve the Fourier-space phase information of the object that speckle interferometry discards. The most popular of these techniques is undoubtedly that proposed by Knox and Thompson (1974) and the following discussion concentrates upon this technique. However, an associated technique due to Aitken *et al.* (1985) is also commented upon.

Like speckle interferometry, Knox-Thompson (KT) is a Fourier-space technique. However, KT explicitly reconstructs the Fourier-space phase information so that the image, rather than just the autocorrelation, can be reconstructed directly. Rather than averaging the spectral intensities, as for speckle interferometry (see (4.13)), thereby throwing away any phase information contained in the speckle images, the complex conjugate of each speckle spectrum is multiplied by a shifted version of that spectrum. An ensemble average described by

$$\begin{aligned} S_{\mathrm{KT}}(\mathbf{u},\boldsymbol{\alpha}) &= \langle S_m^*(\mathbf{u}) S_m(\mathbf{u}+\boldsymbol{\alpha})\rangle_M \\ &= F^*(\mathbf{u})F(\mathbf{u}+\boldsymbol{\alpha})\langle H_m^*(\mathbf{u})H_m(\mathbf{u}+\boldsymbol{\alpha})\rangle_M + \tilde{C}(\mathbf{u}) \end{aligned} \tag{4.17}$$

is then formed, where $\boldsymbol{\alpha}$ is a displacement frequency whose choice is fairly critical in practice. The magnitude of $\boldsymbol{\alpha}$ must be sufficiently large that the ensemble average of LHS of (4.17) is significantly different from $S_{\mathrm{LA}}(\mathbf{u})$. Conversely, it must be sufficiently small that phase correlation is preserved on the average between $S_m(\mathbf{u})$ and $S_m(\mathbf{u}+\boldsymbol{\alpha})$. Note that, when $\boldsymbol{\alpha} = 0$, (4.17) is real and non-negative and reduces to (4.13). When $\boldsymbol{\alpha} \neq 0$, (4.17) is, in general, complex and conveys information about $\mathcal{P}[F(\mathbf{u})]$. The phase of (4.17) yields an estimate of the phase differences in $F(\mathbf{u})$ across a vector distance $\boldsymbol{\alpha}$ in Fourier space, since

$$\begin{aligned} \mathcal{P}[S_{\mathrm{KT}}(\mathbf{u},\boldsymbol{\alpha})] &= \mathcal{P}[F(\mathbf{u}+\boldsymbol{\alpha})] - \mathcal{P}[F(\mathbf{u})] + \\ &\quad \mathcal{P}[\langle H_m^*(\mathbf{u})H_m(\mathbf{u}+\boldsymbol{\alpha})\rangle_M]. \end{aligned} \tag{4.18}$$

The third term on RHS of (4.18) is approximately zero for sufficiently large M (Knox 1976), so (4.18) simplifies to

$$\mathcal{P}\left[S_{\mathrm{KT}}(\mathbf{u},\boldsymbol{\alpha})\right] \approx \mathcal{P}\left[\hat{F}(\mathbf{u}+\boldsymbol{\alpha})\right] - \mathcal{P}\left[\hat{F}(\mathbf{u})\right]. \tag{4.19}$$

The phase of the object, $\mathcal{P}\left[\hat{F}(\mathbf{u})\right]$, is estimated at discrete points in Fourier-space by summing the phase differences along paths radiating from the origin. Since, for real objects, $\mathcal{P}\left[F(0)\right]$ is necessarily zero, $\mathcal{P}\left[\hat{F}(\boldsymbol{\alpha})\right]$ is given immediately by (4.19) when $\mathbf{u}$ is set to zero. Furthermore,

$$\mathcal{P}\left[S_{\mathrm{KT}}(\boldsymbol{\alpha},\boldsymbol{\alpha})\right] = \mathcal{P}\left[\hat{F}(2\boldsymbol{\alpha})\right] - \mathcal{P}\left[\hat{F}(\boldsymbol{\alpha})\right], \tag{4.20}$$

allowing $\mathcal{P}\left[\hat{F}(2\boldsymbol{\alpha})\right]$ to be found because the other quantities in (4.20) are known. In this manner $\mathcal{P}\left[\hat{F}(\mathbf{u})\right]$ is generated recursively for all values for which $|\hat{F}(\mathbf{u})|$ is significant. In one dimension the paths of summation necessarily follow the u-axis. In two dimensions (Knox 1976), each point in Fourier-space can be reached along more than one possible path of summation. Algorithms for following various paths (Knox 1976; Baba *et al.* 1984) and averaging multiple paths (*cf.* Hunt 1979; Frost *et al.* 1979) have been reported. The iterative calculation of $\mathcal{P}\left[\hat{F}(\mathbf{u})\right]$, implicit in the summation of phase differences along the paths, is a type of phase closure (Jennison 1958).

Since $\mathcal{P}\left[\hat{F}(\mathbf{u})\right]$ is derived from a summation of phase differences, accumulated errors adversely affect the accuracy of the phase for large $|\mathbf{u}|$ (Sinton 1986, §3.5.2). Increasing $|\boldsymbol{\alpha}|$ reduces the number of terms to be summed and consequently the potential for error. This must be balanced with the knowledge that choosing too large a value for $|\boldsymbol{\alpha}|$ means that there will be insufficient phase correlation on the average between $S_m(\mathbf{u})$ and $S_m(\mathbf{u}+\boldsymbol{\alpha})$ (Bates 1982b, §8.4). In practice, $|\boldsymbol{\alpha}|$ should be chosen between 0.2 and 0.8 of r_0/D (von der Lühe 1988) (*cf.* Roddier 1988, p. 128). When imaging objects of complicated form, Fontanella and Sève (1987) assert that it is advantageous to adopt a modification to the KT algorithm to accommodate Fourier-space 'phase dislocations' associated with the zeros of $F(\mathbf{u})$.

The usual procedure for retrieving the object involves combining the phase estimate $\mathcal{P}\left[\hat{F}(\mathbf{u})\right]$ obtained from KT processing with the magnitude estimate from speckle interferometry. However, algorithms for determining an image from its Fourier transform phase (see Chapter 5) have also been investigated (Bruck and Sodin 1984). Alternatively, von der Lühe (1988) has proposed models that can calibrate the KT cross spectra provided r_0 is known.

Ayers *et al.* (1988) report that, to achieve maximal SNR from KT processing, it is necessary to centroid the speckle images. They note that difficulties arise in computing the centroid of photon-limited data. They also note that the need to centroid the data increases the computational effort needed to implement KT.

Other techniques based on phase gradient averaging have been proposed, notably by Aitken *et al.* (1985) (*cf.* Aitken *et al.* 1986; Roddier 1988, §4.8), who estimate the phase gradient of $F(\mathbf{u})$, i.e. $\nabla\mathcal{P}\left[F(\mathbf{u})\right]$, directly from the ensemble of speckle images, by forming components of $\langle\nabla\mathcal{P}\left[S_m(\mathbf{u})\right]\rangle_M$. The component for each speckle image is formed by multiplying $s_m(\mathbf{x})$ by $t(\mathbf{x}) = a\mathbf{x} + b$, and invoking the derivative theorem (Table 2.3) of the Fourier transform, where the coefficients a and b are positive constants. These components are then averaged over all speckle images to yield an estimate of the gradient of the object's phase, from which $\mathcal{P}\left[F(\mathbf{u})\right]$ can be reconstructed. Aitken *et al.* (1985) suggest that this formulation allows the processing

to be performed optically, even in real-time, by incorporating a neutral density filter having a linearly graded amplitude transmittance into an optical bench (§2.5). Aitken *et al.* (1986) claim that their technique appears to be less sensitive to bias terms encountered in KT processing. An implementation of phase gradient processing for application with photon-limited data is presented by Johnson and Aitken (1989).

4.7 Triple Correlation

Another Fourier-space astronomical speckle processing method is based on averaging the *(auto)triple correlation* (TC) (Lohmann and Wirnitzer 1984) of the speckle images in the ensemble (Weigelt and Wirnitzer 1983; Lohmann *et al.* 1983). Alternatively, the method can be viewed as averaging the Fourier transform of the triple correlation, called the *bispectrum.* Impressive results have been obtained from TC processing (see, for example, Weigelt and Baier 1985). Triple correlations are finding use in other fields of image processing, for instance pattern recognition (Lohmann 1988) where their shift and rotation invariant properties are utilized (Lohmann 1986; Dainty and Northcott 1986; Bates 1987b).

The TC is a third-order extension of the more familiar second-order operation of autocorrelation. The TC of an arbitrary function $f(\mathbf{x})$ is defined as (Lohmann and Wirnitzer 1984)

$$f^{(\mathrm{TC})}(\mathbf{x}_1, \mathbf{x}_2) = \int\!(K)\!\int f^*(\mathbf{x}) f(\mathbf{x}+\mathbf{x}_1) f(\mathbf{x}+\mathbf{x}_2)\, \mathrm{d}\mathbf{x}. \tag{4.21}$$

Thus, the K-dimensional function produces a $2K$-dimensional triple correlation function. The bispectrum (alternatively called the Erlangen bispectrum (Bates 1987b)) of $f(\mathbf{x})$ is defined by (Lohmann and Wirnitzer 1984)

$$F^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) = F(\mathbf{u}_1) F^*(-\mathbf{u}_2) F^*(\mathbf{u}_1 + \mathbf{u}_2). \tag{4.22}$$

If $f(\mathbf{x})$ is real valued, as it always is in optical astronomy, then $F(\mathbf{u})$ is conjugate-symmetric (i.e. $F(\mathbf{u}) = F^*(-\mathbf{u})$), and (4.22) becomes

$$F^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) = F(\mathbf{u}_1) F(\mathbf{u}_2) F(-\mathbf{u}_1 - \mathbf{u}_2). \tag{4.23}$$

When $f(\mathbf{x})$ is real, eight-fold symmetry occurs in the bispectrum since

$$\begin{aligned} F^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) &= F^{(\mathrm{TC})}(\mathbf{u}_2, \mathbf{u}_1) \\ &= F^{(\mathrm{TC})}(-\mathbf{u}_1 - \mathbf{u}_2, \mathbf{u}_2) \\ &= F^{(\mathrm{TC})*}(-\mathbf{u}_1, -\mathbf{u}_2). \end{aligned} \tag{4.24}$$

The bispectrum of the m^{th} speckle image is defined by

$$\begin{aligned} {S_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) &= F(\mathbf{u}_1) F(\mathbf{u}_2) F(-\mathbf{u}_1 - \mathbf{u}_2) \\ &\qquad H_m(\mathbf{u}_1) H_m(\mathbf{u}_2) H_m(-\mathbf{u}_1 - \mathbf{u}_2) + \tilde{C}(\mathbf{u}) \\ &= F^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) {H_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) + \tilde{C}(\mathbf{u}), \end{aligned} \tag{4.25}$$

i.e. ${S_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2)$ is the product of the bispectra of the object and the speckle psf. TC speckle imaging is often based on the ensemble average of $s_m(\mathbf{x}_1, \mathbf{x}_2)$ (Weigelt and Wirnitzer 1983; Lohmann *et al.* 1983, §III.A), which is

$$\langle s_m(\mathbf{x}_1, \mathbf{x}_2) \rangle_M = f(\mathbf{x}_1, \mathbf{x}_2) \circledcirc \langle h_m(\mathbf{x}_1, \mathbf{x}_2) \rangle_M + \tilde{c}(\mathbf{x}). \tag{4.26}$$

Alternatively, ${S_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2)$ (Lohmann *et al.* 1983, §III.B; Bartelt *et al.* 1984) may be averaged, yielding

$$\begin{aligned} S_{\mathrm{TC}}(\mathbf{u}_1, \mathbf{u}_2) &= \langle {S_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) \rangle_M \\ &= F^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) \langle {H_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) \rangle_M + \tilde{C}(\mathbf{u}). \end{aligned} \tag{4.27}$$

In order to obtain a reconstruction of the object, $\langle {H_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2) \rangle_M$ must be estimated. One way to achieve this is to repeat the triple correlation processing on speckle images of an unresolved reference star (Lohmann *et al.* 1983). Alternative approaches, based on assumptions of the seeing, are available (Lohmann *et al.* 1983). Once the compensation has been performed $\hat{F}^{(\mathrm{TC})}(\mathbf{u})$ can be obtained. Recursive algorithms for obtaining $\mathcal{P}\left[\hat{F}(\mathbf{u})\right]$ from $F^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2)$ are available (Lohmann *et al.* 1983, §III.B; Bartelt *et al.* 1984, §IV.A; Lohmann 1986).

Alternatively, $\mathcal{P}\left[\hat{F}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2)\right]$ can be obtained directly from $s_{\mathrm{TC}}(\mathbf{u}_1, \mathbf{u}_2)$, without needing to compensate for seeing, since ${H_m}^{(\mathrm{TC})}(\mathbf{u}_1, \mathbf{u}_2)$ is real-valued (Lohmann *et al.* 1983, Appendix B; Bates 1987b). Thus, an estimate of the true image can be formed by combining this estimate of $\mathcal{P}\left[\hat{F}(\mathbf{u})\right]$ with $|\hat{F}(\mathbf{u})|$ generated by, for example, speckle interferometry, and taking the inverse Fourier transform.

The effect of noise on TC analysis has been studied. Wirnitzer (1985) analyses of the effects of photon limiting on TC processing. He concludes that image reconstruction is possible for all cases in which speckle interferometry is successful, provided the average number of photons per speckle image exceeds four. Nakajima (1988) derives a general expression for the SNR of the bispectrum at arbitrary light levels. Northcott *et al.* (1988) describe three algorithms for implementing TC for data which are severely photon-limited. They report that a reconstruction based on the projection theorem, although not giving quite as faithful reconstructions as a direct algorithm, is superior from an implementational viewpoint. Roddier (1988, §4.9) notes that TC appears to be slightly more robust than Knox-Thompson in the presence of contamination.

Implementation of TC tends to be computationally expensive because of the need to store and average four-dimensional data. However, Granrath (1987) is investigating the feasibility of implementing an optimized TC algorithm on a highly parallel computer at video rates.

Hofmann and Weigelt (1987) have recently reported a new algorithm based on the cross triple correlation. The advantage of this method stems from undesired photon bias terms in (4.27) being eliminated. They report computer simulations showing the apparent feasibility of the method.

In a recent comprehensive comparison of the Knox-Thompson and TC methods, Ayers *et al.* (1988) report that these two correlation techniques are closely related, both from a theoretical and an implementational point of view. They note that by allowing the vector $\boldsymbol{\alpha}$ appearing in the analysis of KT processing (see (4.17)) to assume several different values, KT utilizes redundancy of phase information that effectively increases the SNR. Under these conditions they conclude that the SNR of KT and TC are similar. They also interpret speckle interferometry, KT and TC as examples of phase closure.

4.8 Shift-and-add

The shift-and-add (SAA) algorithm (Bates 1976; Bates and Cady 1980) is the most conceptually and computationally simple of all proposed astronomical speckle imag-

ing algorithms. This algorithm is discussed in more detail than the speckle imaging algorithms presented in previous sections, since many of the examples, which are presented in Chapter 7 to illustrate general blind deconvolution algorithms, involve improving the images resulting from SAA.

The basic SAA algorithm is discussed in §4.8.1. This technique only gives faithful restorations for a restricted class of objects — those possessing a dominatingly bright unresolvable feature. When an object does not possess such a feature, the SAA image exhibits artefacts, called *ghosts*. The mechanism of ghost formation is outlined in §4.8.2. Extensions to basic SAA, presented in §4.8.3, enable a broader range of true images to be processed.

SAA was initially developed for use with astronomical speckle data. However, the technique is proving useful in other fields, notably ultrasonic imaging and speech processing. Applications of suitably modified forms of the SAA algorithm to these fields are discussed in §4.8.4.

A major impetus for my pursuing algorithms for the general blind deconvolution problem (§1.5) has been the ghosting problem of basic SAA. Algorithms for implementing general blind deconvolution are described in Chapter 7. These algorithms have proven effective in 'deghosting' SAA images (Bates and Davey 1987b; Bates and Davey 1988), although they have much wider potential application.

The connection between deghosting and blind deconvolution arises because the image resulting from SAA is expressible as a convolution of the object with an object-dependent psf (see §4.8.1). Thus, a general blind deconvolution algorithm could deconvolve these two components, thereby deghosting the SAA image. Combining the basic SAA algorithm with such a technique would allow most of the speckle processing to be implemented by the computationally efficient basic SAA algorithm. Then only a single, more computer intensive, operation would be required to deghost the final image. Such a technique would have a distinct advantage over the extensions to SAA discussed in §4.8.3, since each of these extensions detracts from the simplicity of basic SAA.

4.8.1 Basic Shift-and-add

Basic SAA (Bates 1976) consists of the following steps. The brightest pixel in each speckle image in the ensemble is located. Each speckle image is then shifted, without rotation, so that its brightest pixel lies at the centre of the image. All the shifted images are summed. The *SAA image*, $f_{\mathrm{sa}M}(\mathbf{x})$, is then formed by dividing each pixel in the resulting image by M. Thus, the basic SAA operation is described by

$$\begin{aligned} f_{\mathrm{sa},M}(\mathbf{x}) &= \langle s_m(\mathbf{x}+\mathbf{x}_m)\rangle_M \\ &= \langle f(\mathbf{x}) \odot h_m(\mathbf{x}+\mathbf{x}_m)\rangle_M + \tilde{c}(\mathbf{x}) \\ &= f(\mathbf{x}) \odot h_{\mathrm{sa}}(\mathbf{x}) + \tilde{c}(\mathbf{x}), \end{aligned} \tag{4.28}$$

where $\mathbf{x}_m$ is the coordinate of the brightest pixel in the m^{th} recorded image and

$$h_{\mathrm{sa}}(\mathbf{x}) = \langle h_m(\mathbf{x}+\mathbf{x}_m)\rangle_M \tag{4.29}$$

is the *SAA psf*. Equation (4.28) reveals that the SAA image can be considered to be the contaminated convolution of the object with the SAA psf.

Since SAA is, in effect, a method of registering individual speckle images, it makes sense to call the brightest pixel in any particular speckle image (i.e. $\mathbf{x}_m$ in (4.28)) the *SAA reference* for that speckle image. It is also convenient to call the brightest pixel in $f(\mathbf{x})$ the *object reference*. The underlying idea of SAA is that

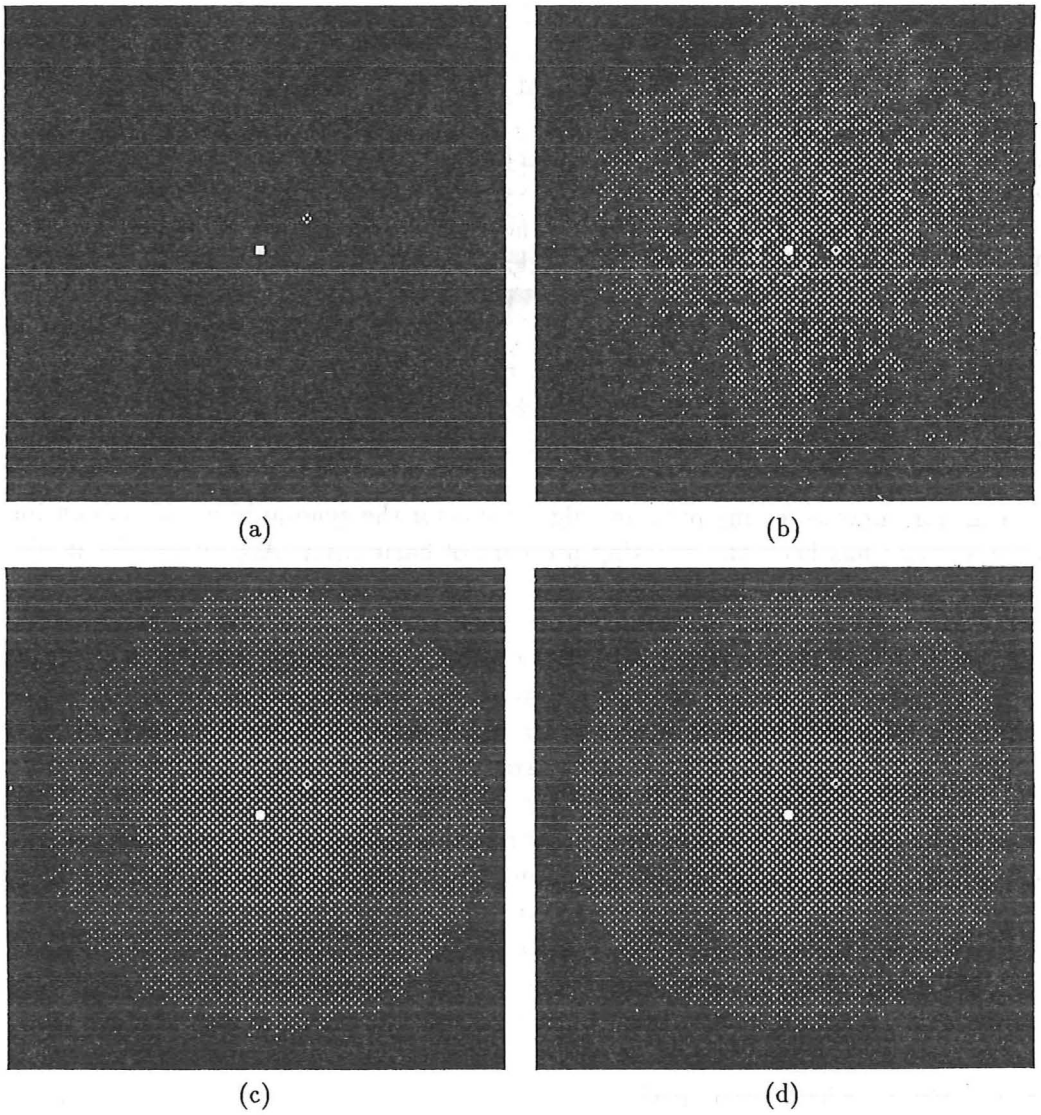
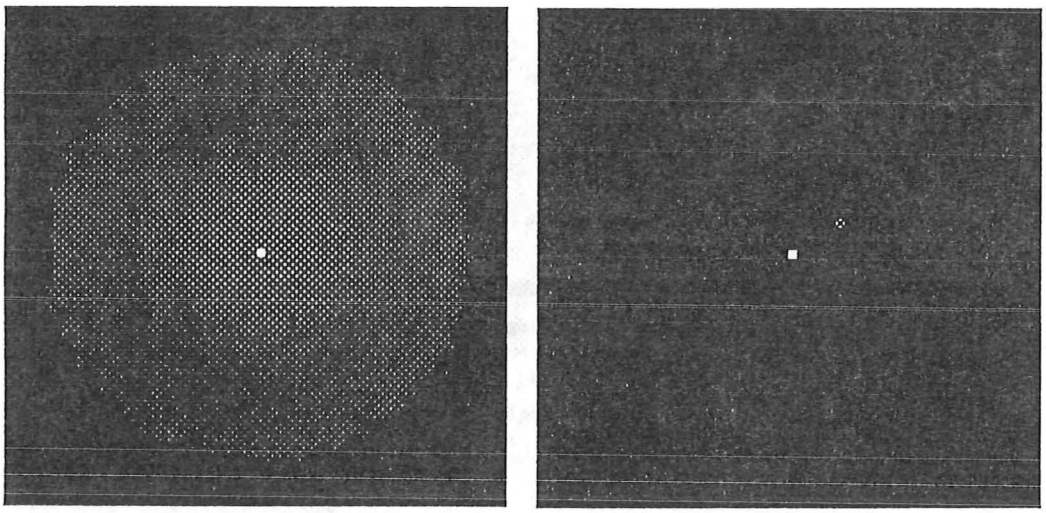


Figure 4.6: Computer simulation of basic SAA applied to speckle images of a binary star ($R = 0.3$) which were contaminated to a level of $\epsilon_{\mathrm{dB}} = -30\mathrm{dB}$: (a) $f(x, y)$ for this simulation; (b) $f_{\mathrm{sa},4}(x, y]$); (c) $f_{\mathrm{sa},64}(x, y)$; (d) $f_{\mathrm{sa},256}(x, y)$.

a blurred image is likely to be least distorted where it has the greatest magnitude (Bates 1976). Thus, by averaging the speckle images that have been shifted so their brightest pixel lies at the image centre, one is effectively averaging the least distorted of the randomly weighted and positioned versions of the object that comprise the speckle image (see §4.2).

The versions of the object, whose object references were not chosen to be the SAA reference, average to produce a *fog*. The fog becomes smoother as more speckle images are processed, until, when M is large, it is similar in form to the seeing disc. The SAA image then consists of the reconstruction of the object sitting atop a fog.

Fig. 4.6 illustrates basic SAA being applied to computer generated speckle images of the binary star depicted in Fig. 4.6(a). Two hundred and fifty six contaminated speckle images of this object were generated in the manner described in §4.4.2, and illustrated by Fig. 4.5. The basic SAA image resulting from processing all these images is shown in Fig. 4.6(d). Inspection of this reconstruction reveals that, as



(a) (b)

Figure 4.7: Defogging the SAA image shown in Fig. 4.6(d) in the manner described by (4.31): (a) $f_{\delta,\mathrm{sa},256}(\mathbf{x})$; (b) $\hat{f}(\mathbf{x})$.

expected, the basic SAA image comprises two peaks sitting on a smooth fog.

The fog can be removed from the SAA image with the aid of the SAA image of a single unresolved star viewed under seeing conditions statistically similar to those applying when the object of interest was observed. This image can be written

$$\begin{aligned} f_{\delta,\mathrm{sa},M_\delta}(\mathbf{x}) &= \langle s_{\delta,m}(\mathbf{x}+\mathbf{x}_m)\rangle_{M_\delta} \\ &= \langle f_\delta(\mathbf{x}) \odot h_m(\mathbf{x})\rangle_{M_\delta} \\ &= a\langle h_m(\mathbf{x})\rangle_{M_\delta}, \end{aligned} \tag{4.30}$$

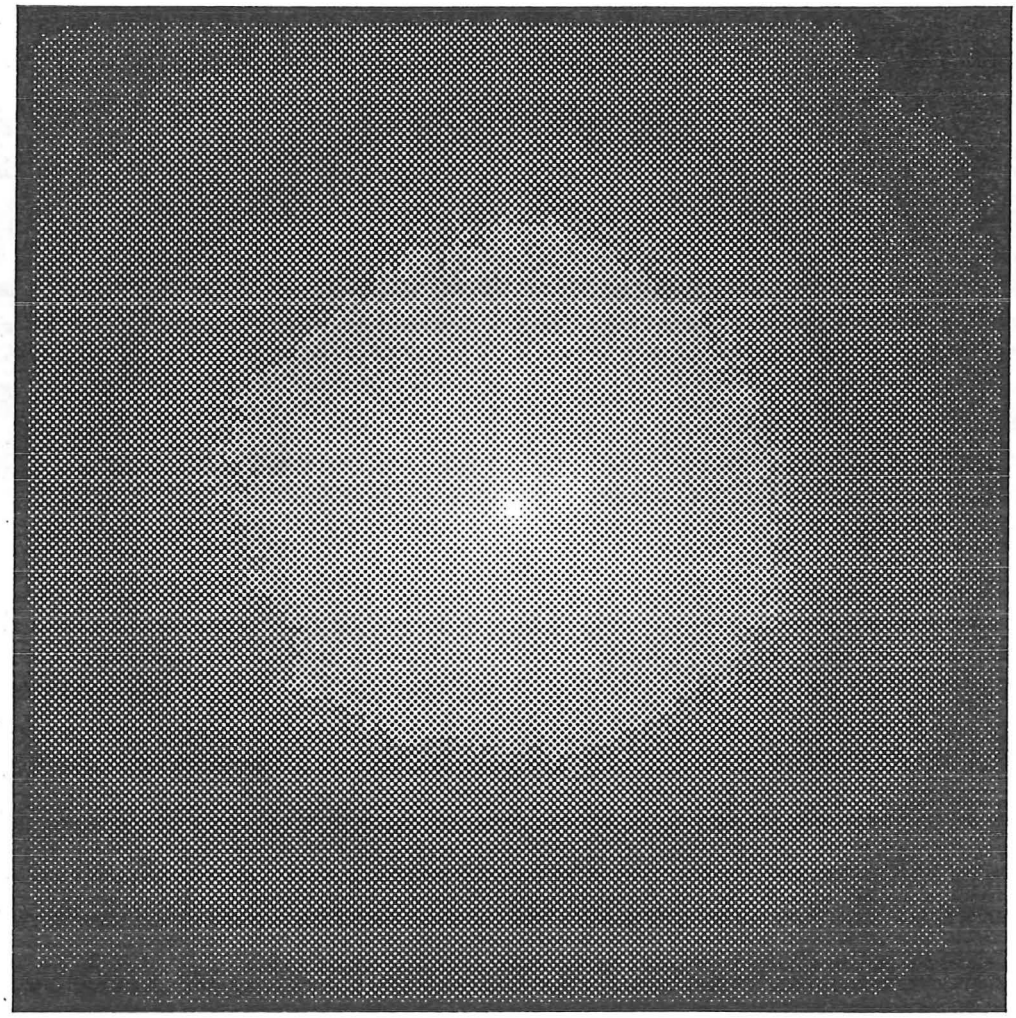
where the subscript δ signifies the reference star. A Wiener filter (§3.1.2) derived from $f_{\delta,\mathrm{sa},M_\delta}(\mathbf{x})$ allows a 'defogged' estimate of the object i.e.

$$\hat{f}(\mathbf{x}) = f_{\mathrm{sa},M}(\mathbf{x}) \odot^{-1} f_{\delta,\mathrm{sa},M_\delta}(\mathbf{x}) \tag{4.31}$$

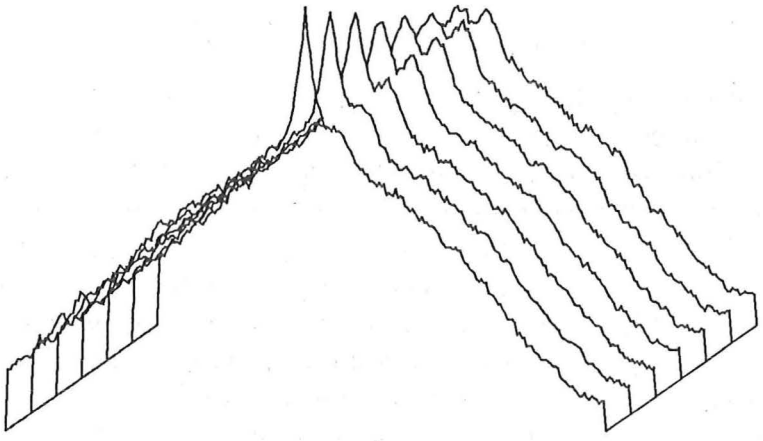
to be recovered (see Fig. 4.7). The symbol $\odot^{-1}$ appearing in (4.31) signifies the deconvolution achieved by Wiener filtering the quantity preceding the symbol with a filter derived from the quantity which follows. Other methods of removing the fog, without recourse to a reference object, have been reported (Sinton *et al.* 1986; *cf.* Baba *et al.* 1987; Christou *et al.* 1985).

Basic SAA has been shown to contribute useful information about the positioning of components of multiple stars even when operating upon wide bandwidth speckle images. Bates and Cady (1980) report optical laboratory simulations in which speckle images having a bandwidth of 100nm are recorded. SAA applied to these images reveals the correct juxtaposition of the components of the star, although with somewhat lower contrast than is achieved with narrow bandwidth speckle images. Baba *et al.* (1985) apply SAA to wide bandwidth ($\Delta\lambda$ = 300nm) speckle images of the binary α Com. The resulting SAA image shows the components clearly resolved. The result of applying SAA to 86 wide bandwidth ($\Delta\lambda$ = 300nm) speckle images of the star ADS 2200 is shown in Fig. 4.8(a). The accompanying plots, Fig. 4.8(b), confirm that SAA has clearly resolved the secondary peak of the star.

SAA has also been applied to photon-limited speckle images. The most obvious feature of a SAA image generated from such data is a narrow spike on top of the peak



(a)



(b)

Figure 4.8: SAA applied to wide bandwidth speckle images of the star ADS 2200. One of the speckle images in the ensemble is shown in Fig. 4.3: (a) $f_{\mathrm{sa},86}(\mathbf{x})$; (b) line plots showing the intensity of contiguous rows of (a). The bottommost plot corresponds to the row containing the image centre. Successively higher plots correspond to successively higher rows. The secondary peak of the star is clearly resolved in this SAA image despite the very wide bandwidth used for the recording. ($D = 1.88$m, $\lambda_0 \approx 450$nm, $\Delta\lambda = 300$nm, $\tau_s \approx 8$ms.) Speckle images courtesy of University of Hokkaido.

at the centre of the SAA image. Following Christou *et al.* (1985) this is called the *photon spike*. Christou *et al.* (1985; 1986b) have demonstrated that, in practice, the photon spike can be removed from the SAA image if a model of the psf characterizing the apparatus which made the recording is available. The removal is effected by fitting the spectrum of the model to the spectrum of the SAA image beyond the diffraction limit of the telescope. The fitted model is then subtracted from the SAA image's spectrum to yield the spectrum of the 'despiked' SAA image.

Sinton (1986, §5.4) reports one-dimensional computer simulations of basic SAA imaging applied to speckle images of a binary star under varying degrees of photon limiting. He concludes that, when the average number $\bar{N}_p$ of photons per speckle image is greater than 1000, virtually no degradation is apparent in the reconstruction compared with that formed in non-photon-limited conditions. However, when $\bar{N}_p$ is decreased further, degradation becomes apparent in the reconstruction although the binary is still clearly resolved, even when $\bar{N}_p = 5$. Under conditions of such severe photon limiting the SAA image resembles the autocorrelation of the object. This reveals that SAA has preserved less of the Fourier-space phase information, and has thus lost much of its advantage over speckle interferometry.

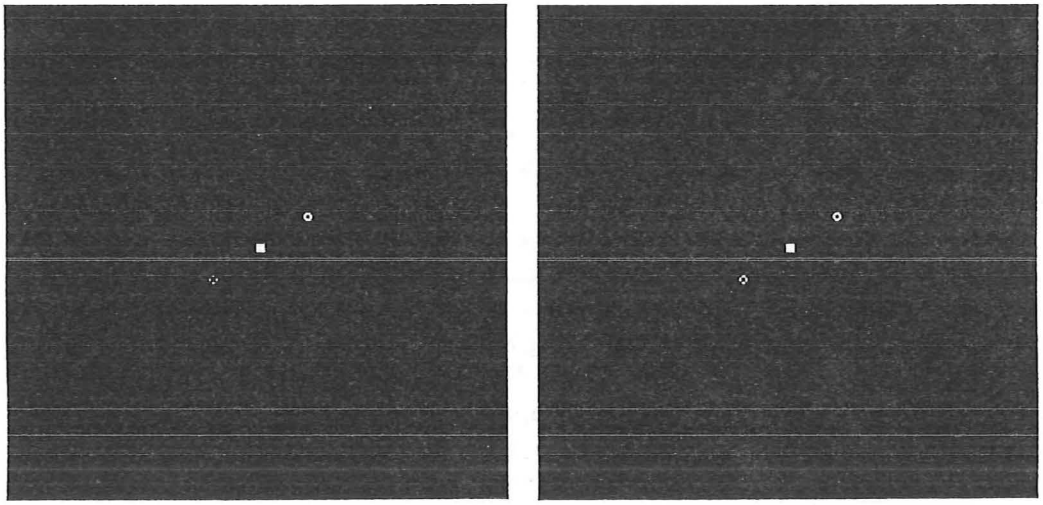
4.8.2 Shortcomings of Basic Shift-and-add

The basic SAA image can only be faithful if the same point in the object always acts as the SAA reference in each speckle image. Thus, SAA generates faithful images without any further processing when the object has a single pixel with a much greater amplitude than the remainder of the object, i.e. it has a well defined object reference. For speckle images of this class of object, the object reference in the brightest version of the object provides a reliable SAA reference.

For speckle images of objects that are not dominated by a single brightest pixel, the SAA reference may correspond to any of the brighter pixels in the object. This ambiguity implies that the brightest versions of the object present in some speckle images do not occupy the same position in the SAA image as the brightest versions present in other speckle images. This leads to artefacts, called *ghosts* (Bates and Cady 1980; Cady 1980), in the resulting SAA image (see Fig. 4.9).

The SAA image of a binary star, in general, exhibits three peaks. This is clearly shown in Fig. 4.9(a) and (b) in which the brightest peak, the *primary peak*, lies at the image centre. The next brightest peak, the *secondary peak*, appears at the true position of the secondary component of the binary, i.e. at a vector spacing $\mathbf{s}$ from the primary peak. The smallest peak, the *ghost peak*, is found at a vector spacing $-\mathbf{s}$ from the primary peak. It is convenient to denote the brightnesses of these peaks in the reconstruction as I_p, I_s and I_g respectively.

A graph of the computer simulated values of I_p and I_s versus the true value of R is shown in Fig. 4.10 for a contamination level of $\epsilon_{dB} = -30$dB added to the speckle images. This data was compiled from a sequence of basic SAA images of simulated binary stars, having the same separation as the binary depicted in Fig. 4.6(a), but with values of R ranging from 0.0 to 1.0 in steps of 0.05. The basic SAA images of these stars were generated in the same manner as those shown in Fig. 4.7 and Fig. 4.9(a) and (b). Inspection of Fig. 4.10 reveals that, as expected, the ghost peak becomes increasingly brighter as R increases. It is also apparent that an increase in I_g is coupled with a decrease in I_s suggesting that energy from the secondary peak forms the ghost peak. An ideal algorithm would produce no ghosting, so the dashed lines would lie along the horizontal axis. An ideal algorithm would also reconstruct I_s exactly, so the solid lines would superimpose upon the dotted line. Departures of



(a) (b)

Figure 4.9: Ghosting in SAA images of binary stars. These 'defogged' SAA images were generated in the same manner as the image depicted in Fig. 4.7, but with the object being a binary star of brightness ratio: (a) $R = 0.7$ (b) $R = 0.95$.

the dotted and solid lines from their ideal values represent limitations of basic SAA.

Bagnuolo, Jr. (1982) and Hunt *et al.* (1983, §4) present analytical descriptions of SAA ghosts for objects comprising unresolved parts (e.g. binary stars). Hunt *et al.* (1983, §5) and Bagnuolo, Jr. (1985) discuss ghosting of objects which have extended parts. The extensions to SAA that are discussed in §4.8.3 all aim to reduce the ghosting problem inherent in basic SAA.

It follows from (4.28) that, even in the presence of ghosting, the SAA image can be considered to be the contaminated convolution of the object with the SAA psf. Thus, the ghosting must be incorporated in this psf. Since the ghosting is dependent upon the form of the object, it follows that the SAA psf must also be object dependent. Bates and Davey (1987b) demonstrate that it is often appropriate to separate the seeing- and object-dependent parts of the SAA psf. They show that it is useful to consider that two functions describing these parts can be defined, which, when convolved, yield the SAA psf. Thus, if $h_{\mathrm{sa}_s}(\mathbf{x})$ and $h_{\mathrm{sa}_o}(\mathbf{x})$ denote the seeing-dependent and object-dependent parts of $h_{\mathrm{sa}}(\mathbf{x})$,

$$h_{\mathrm{sa}}(\mathbf{x}) \approx h_{\mathrm{sa}_s}(\mathbf{x}) \odot h_{\mathrm{sa}_o}(\mathbf{x}). \tag{4.32}$$

Consequently, (4.28), the equation defining basic SAA, can be rewritten as

$$f_{\mathrm{sa},M}(\mathbf{x}) = f(\mathbf{x}) \odot h_{\mathrm{sa}_s}(\mathbf{x}) \odot h_{\mathrm{sa}_o}(\mathbf{x}) + \tilde{c}(\mathbf{x}), \tag{4.33}$$

where the contamination function $\tilde{c}(\mathbf{x})$ again incorporates any discrepancies between the model and actuality.

As discussed in §4.8.1, an estimate of the fog can be obtained by applying SAA to speckle images of an unresolved reference star. Since in this situation the object is unresolved, it follows that the resulting image provides an estimate of the seeing dependent component of the SAA psf. Thus, the image defined by (4.30) yields an estimate of $h_{\mathrm{sa}_s}(\mathbf{x})$, i.e.

$$h_{\mathrm{sa}_s}(\mathbf{x}) \approx f_{\delta,\mathrm{sa},M_\delta}(\mathbf{x}). \tag{4.34}$$

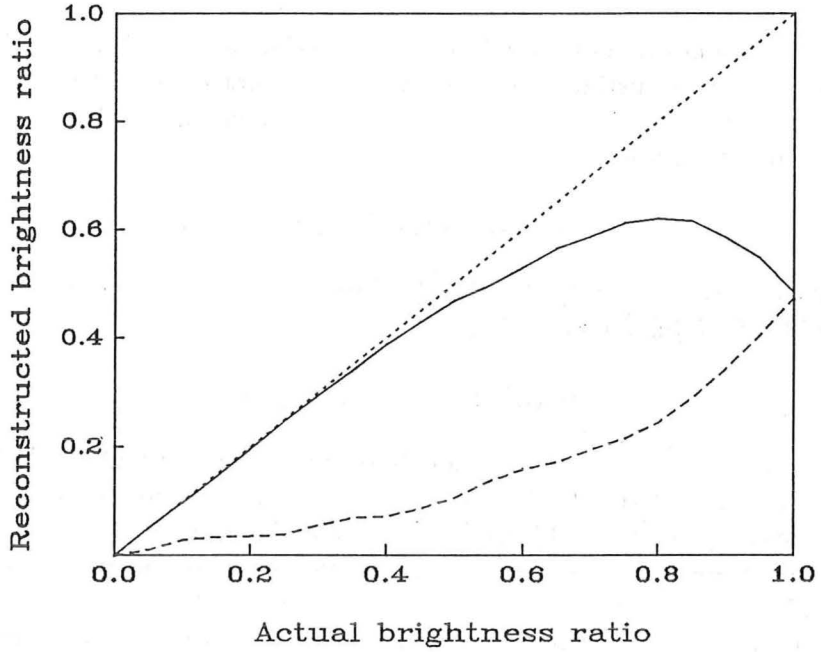


Figure 4.10: The dependence of I_s (solid line) and I_g (dashed line) upon R. This data was compiled from a sequence of basic SAA images of a binary star generated in the same manner as those presented in Fig. 4.9(a) and (b). Values of R span from 0.0 to 1.0 increments of 0.05, and the data values are normalized such that $I_p = 1.0$.

By combining (4.33) and (4.34) it follows that

$$\begin{aligned} \tilde{f}_{\mathrm{sa},M}(\mathbf{x}) &= f_{\mathrm{sa},M}(\mathbf{x}) \odot^{-1} f_{\delta,\mathrm{sa},M_\delta}(\mathbf{x}) \\ &\approx f(\mathbf{x}) \odot h_{\mathrm{sa}_o}(\mathbf{x}), \end{aligned} \tag{4.35}$$

where $\tilde{f}_{\mathrm{sa}}(\mathbf{x})$ denotes the SAA image with the seeing dependent component of the SAA psf removed. This relationship can be exploited when applying general deconvolution algorithms to deghost the SAA image (Bates and Davey 1987b). Examples of such processing are also provided in Chapter 7.

4.8.3 Extensions to Basic Shift-and-add

Several extensions to basic SAA have been proposed with the aim of increasing the faithfulness of SAA images of particular classes of object. The processing is significantly more intricate for each of these extensions than for basic SAA. It is interesting to note that, although the processing becomes more complicated, the comparative insensitivity to contamination is preserved because it is still the original speckle images that are shift-and-added.

The first extension to SAA that is discussed here is an extension in concept only, since it was actually proposed prior to SAA and can be considered as that technique's forerunner. This method, here abbreviated to LWH, was proposed and demonstrated by Lynds *et al.* (1976) (*cf.* Sinton 1986, §3.6.1). In the method, all the speckles having a brightness above some threshold are averaged. This is achieved by generating, for each speckle image, a mask of weighted impulses representing the positions and brightnesses of the individual speckles in $h_m(\mathbf{x})$. The m^{th} such mask is defined by

$$m_m(\mathbf{x}) = \sum_{n=1}^{N_m} |s_m(\mathbf{x}_{m,n})| \delta(\mathbf{x} - \mathbf{x}_{m,n}), \tag{4.36}$$

where $\mathbf{x}_{m,n}$ is the position of the n^{th} brightest speckle in the m^{th} speckle image and N_m is the number of distinct brightest speckles in that image. The mask formed from each speckle image is correlated with its corresponding speckle image to form a modified image defined by

$$s'_m(\mathbf{x}) = m_m(\mathbf{x}) \star s_m(\mathbf{x}), \tag{4.37}$$

where $\star$ denotes correlation (§2.2). The final image is obtained by applying SAA to the ensemble of modified images, i.e.

$$f_{\text{LWH}}(\mathbf{x}) = \langle s'_m(\mathbf{x} + \mathbf{x}'_m) \rangle_M, \tag{4.38}$$

where $\mathbf{x}'$ is the brightest pixel in $s'_m(\mathbf{x})$. Bates and Milner (1978) (*cf.* Milner 1979) describe an extension to the LWH method that allows objects consisting of separated, unresolvable components (*cf.* (4.5)) to be faithfully imaged.

The *correlation extension* (Minard *et al.* 1985) (*cf.* Bates *et al.* 1985; Sinton *et al.* 1986) to SAA (CAA) is an iterative procedure that aims to locate the SAA reference positions more accurately, thereby reducing ghosting. This is achieved by generating an estimate of the object, $\hat{f}(\mathbf{x})$, either from *a priori* information or, more commonly, from the output of basic SAA. $\hat{f}(\mathbf{x})$ is correlated with each speckle image in turn, with the brightest point of the modified images being taken as the SAA reference point for the corresponding original speckle image (Minard 1985). Thus,

$$s'_m(\mathbf{x}) = \hat{f}(\mathbf{x}) \star s_m(\mathbf{x}), \tag{4.39}$$

and

$$f_{\text{ca},M}(\mathbf{x}) = \langle s'(\mathbf{x} + \mathbf{x}'_m) \rangle_M, \tag{4.40}$$

where $f_{\text{ca}}(\mathbf{x})$ is the *correlation SAA image* and $\mathbf{x}'_m$ is the brightest pixel in the modified image $s'_m(\mathbf{x})$. An improved estimate of the object can be obtained from $f_{\text{ca},M}(\mathbf{x})$ and the processing described by (4.39) and (4.40) repeated. Since the updated estimate of the object is expected to be more faithful, one would expect the SAA reference in each image to be located with more accuracy. Thus, an iterative loop can be formed with successively more faithful estimates of the object being used in the correlation (4.39) (Sinton 1986, §3.6.3). Processing ceases when successive versions of $f_{\text{ca},M}(\mathbf{x})$ differ by less than a prescribed threshold. The correlation SAA image exhibits significantly less ghosting than the basic SAA image formed from the same speckle images (Minard *et al.* 1985) (see Fig. 4.11). Again, the fog may be removed by Wiener filtering $f_{\text{ca},M}(\mathbf{x})$ with $f_{\delta,\text{sa},M_\delta}(\mathbf{x})$.

Ribak *et al.* (1985) independently developed a technique that is similar to correlation SAA. Here, a matched filter,

$$\hat{g}(\mathbf{x}) = \hat{f}^*(-\mathbf{x}), \tag{4.41}$$

is convolved with each speckle image in the ensemble. The brightest pixel in this modified image becomes the SAA reference for the corresponding original speckle image. In the same way as for correlation SAA (Minard *et al.* 1985), an iterative loop can be developed with the matched filter successively being updated as more faithful reconstructions become available. The matched filter improves the accuracy with which speckles can be detected, thereby reducing ghosting.

Baba *et al.* (1987) describe a method for iteratively updating the matched filter when a binary star is being observed. They note that, since a binary is the sum of the

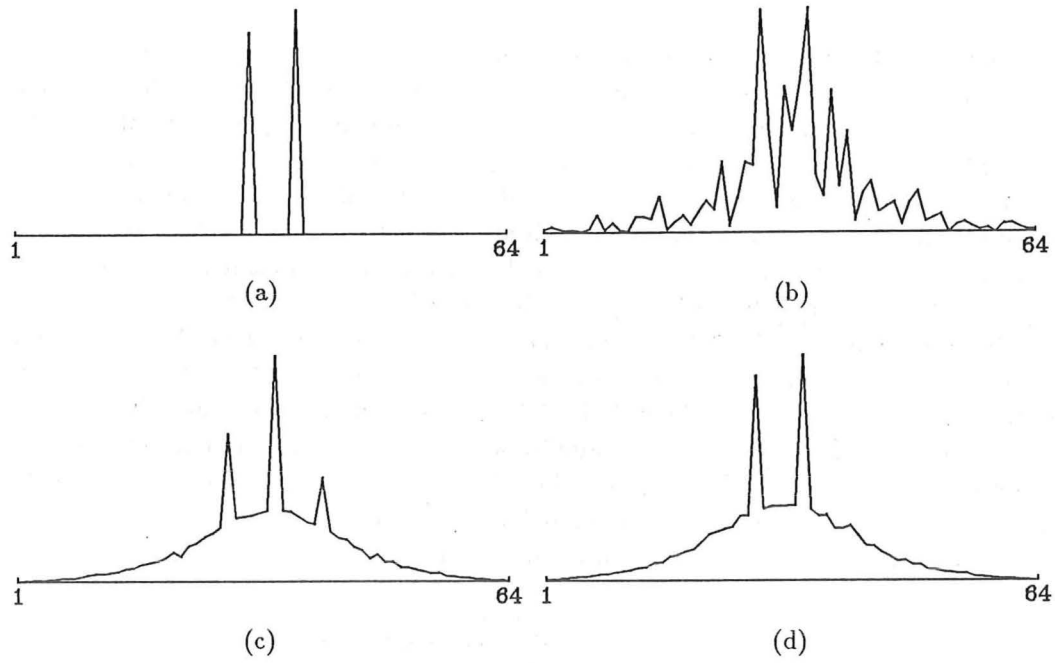


Figure 4.11: A one-dimensional computer simulation of correlation SAA. The speckle images were generated in the manner described in §4.4.1: (a) $f(\mathbf{x})$; (b) a typical $s_m(\mathbf{x})$; (c) $f_{\mathrm{sa},256}(\mathbf{x})$; (d) $f_{\mathrm{ca},256}(\mathbf{x})$ after five iterations of the correlation SAA algorithm. It is readily apparent that the basic SAA image exhibits significantly more ghosting than the correlation SAA image.

speckle psf with a shifted and weighted version of itself (see (4.7)), the relationship between I_{p}, I_{s} and I_{g} may be written

$$I_{\mathrm{p}} = RI_{\mathrm{s}} + I_{\mathrm{g}}/R, \tag{4.42}$$

which is a quadratic in R. In the absence of ghosting, $I_{\mathrm{g}} = 0$, so R can readily be estimated as

$$R = \frac{I_{\mathrm{p}}}{I_{\mathrm{s}}}. \tag{4.43}$$

However, if the ghosting is not negligible, $I_{\mathrm{g}} \neq 0$ and it is necessary to solve (4.42) for the variable R. This yields the relationship

$$R = \frac{I_{\mathrm{p}} - \left(I_{\mathrm{p}}^2 - 4I_{\mathrm{g}}I_{\mathrm{s}}\right)^{1/2}}{2I_{\mathrm{g}}}. \tag{4.44}$$

An estimate of R computed in this manner allows a matched filter, defined by

$$\hat{g}(\mathbf{x}) = a_d(\mathbf{x}) + Ra_d(\mathbf{x} - \mathbf{s}), \tag{4.45}$$

to be formed, where $a_d(\mathbf{x})$ is an estimate of the psf characterizing the apparatus which recorded the speckle image. This matched filter is then used in the manner of (4.39) and (4.40) to accurately locate the SAA reference of each speckle image when SAA is reapplied to the speckle images. Successively more faithful values of R are obtained by iteratively repeating this processing. Iterations cease when the differences between two successive values of R are sufficiently small according to a pre-set criterion, or after a pre-set number of iterations.

The method of Baba *et al.* also has the advantage that observations of a reference star are not required to remove the fog from the SAA image. They effect fog removal by extracting an estimate of the seeing dependent component of the SAA psf directly from the SAA image. The estimate is then subtracted from the original image to produce the defogged image. Baba *et al.* (1987) report that, when processing computer simulated speckle images of a binary star with their extension to SAA, a significantly more accurate value for R was obtained than from basic SAA. They also report applying their technique to speckle images of the binary σ And.

Christou *et al.* (1985) (*cf.* Freeman *et al.* 1985; Christou *et al.* 1986b; 1986a; Ribak 1986) describe a modification to SAA, called *weighted SAA*, which is a close relative of LWH processing. Weighted SAA aims to produce an image that is representative of the true image, whilst simultaneously eliminating the background fog. As for LWH processing, a mask of weighted impulses is generated for each speckle image, the mask for the m^{th} such image being defined by (4.36). The weighted SAA image $f_{\text{WSA}}(\mathbf{x}) \longleftrightarrow F_{\text{WSA}}(\mathbf{u})$ is then computed as

$$\begin{aligned} F_{\text{WSA}}(\mathbf{u}) &= \frac{\langle \dfrac{S_m(\mathbf{u})}{M_m(\mathbf{u})} |M_m(\mathbf{u})|^2 \rangle_M}{\langle |M_m(\mathbf{u})|^2 \rangle_M} \\ &= \frac{\langle S_m(\mathbf{u}) M_m^*(\mathbf{u}) \rangle_M}{\langle M_m(\mathbf{u}) M_m^*(\mathbf{u}) \rangle_M}. \end{aligned} \tag{4.46}$$

The inverse Fourier transform of (4.46) yields

$$f_{\text{WSA}}(\mathbf{x}) = \langle m_m(\mathbf{x}) \star s_m(\mathbf{x}) \rangle_M \odot^{-1} \langle m_m(\mathbf{x}) \star m_m(\mathbf{x}) \rangle_M, \tag{4.47}$$

where $\odot^{-1}$ represents the deconvolution achieved by dividing by $\langle M_m(\mathbf{u}) M_m^*(\mathbf{u}) \rangle_M$ in Fourier-space. The first term on the RHS of (4.47) corresponds to $f_{\text{LWH}}(\mathbf{x})$, while the second term provides an estimate of the fog. Thus, the deconvolution contained in (4.47) automatically removes an estimate of the fog. The efficacy of weighted SAA in removing fog is apparent in images presented by Christou *et al.* (1986b).

Another astronomical speckle imaging algorithm, which can be applied to speckle images of binary stars, is the *Fork algorithm* (Bagnuolo, Jr. 1988). Although not strictly an extension of SAA, this technique is included here since it is an image-space technique having similarities to SAA.

The essence of the Fork algorithm is to search each speckle image for all isolated pairs of speckles that correspond to a replica of the binary. Suppose that I_1, I_2, I_3 and I_4 are the observed intensities at points in the speckle image separated, one from the next, by $\mathbf{s}$. Equation (4.7) reveals that the speckle image of a binary star can be considered to be the speckle psf added to a translated and weighted version of itself. Consider five pixels that are spaced consecutively by $\mathbf{s}$ in the speckle psf. Denoting these as i_0, i_1, i_2, i_3 and i_4, it follows that

$$\begin{aligned} I_1 &= i_1 + R i_0, \qquad I_2 = i_2 + R i_1, \\ I_3 &= i_3 + R i_2, \qquad I_4 = i_4 + R i_3. \end{aligned} \tag{4.48}$$

In general, it is not possible to solve for R from (4.48) since there are too many unknowns. However, suppose that, by chance, i_2 is an isolated speckle, by which it is meant that $i_2 \gg i_1$ and $i_2 \gg i_3$. In this situation I_1 and I_4 are negligible, so I_2 and I_3 form a nearly isolated pair. Consequently, it is possible to obtain an estimate of R:

$$R \approx I_3 / I_2. \tag{4.49}$$

The Fork algorithm selects such 'favourable occurrences' by requiring

$$\begin{aligned} \max(I_2, I_3) &> c_1 \max(I_1, I_4) \\ &> c_2 \bar{I}, \end{aligned} \tag{4.50}$$

where c_1 and c_2 are chosen constants, $\bar{I}$ is the average intensity at the centre of the speckle image, and max is here to be interpreted as 'the greater of'. The second condition applies when significant contamination is present. An improved estimate of the intensity ratio is given by subtracting an estimate of the background $B = (I_1 + I_4)/2$, yielding

$$R \approx \frac{I_3 - B}{I_2 - B}. \tag{4.51}$$

An even more faithful estimate of R is obtained by compensating for the terms included in I_2 and I_3, due to i_2 not being perfectly isolated. Inspection of (4.48) reveals that these are Ri_1 and i_3. Bagnuolo, Jr. (1988) shows that it is possible to estimate these terms if one assumes that the probability intensity i_n can be approximated by an exponential distribution. Upon defining the equations

$$\begin{aligned} \bar{i}_1 &= I_1 R_1 - \frac{1}{Q} \\ \bar{i}_4 &= I_4 R_4 - \frac{1}{Q} \\ \bar{i}_3 &= \frac{\bar{i}_4}{R} \\ \text{where} \quad Q &= \frac{1}{R} - 1 \\ \text{and} \quad R_n &= \frac{e^{in}Q}{e^{inQ} - 1}, \end{aligned} \tag{4.52}$$

it follows that the corrected estimate of the intensity ratio for this occurrence is given by

$$R \approx \frac{I_3 - \bar{i}_3}{I_2 - R\bar{i}_1} \equiv \frac{b}{a}, \tag{4.53}$$

where a and b are positive real variables.

Bagnuolo, Jr. (1988) also suggests that a and b should be weighted by their estimated uncertainties. On denoting uncertainties in $\bar{i}_1$ (and $\bar{i}_3$ similarly) by $\Delta i_1^2 = \bar{i_1^2} - \bar{i_1}^2$, and after defining α to be the fraction of the intensity in the lesser component (i.e. $\alpha = R/(1+R)$), one obtains

$$\begin{aligned} \Delta\alpha^2 &= \frac{b^2\Delta a^2 + a^2\Delta b^2}{(a+b)^4}, \\ \text{where} \quad \Delta a^2 &= R^2\left[I_1^2(1-R_1)R_1 + \frac{1}{Q^2}\right] = \Delta i_1^2 R^2 \\ \text{and} \quad \Delta b^2 &= \left(\frac{1}{R^2}\right)\left[I_4^2(1-R_4)R_4 + \frac{1}{Q^2}\right] = \frac{\Delta i_4^2}{R^2}. \end{aligned} \tag{4.54}$$

Thus, one can compute the value of α for each favourable occurrence, weight it by its corresponding $\Delta\alpha$, and store it in a histogram. After all speckle images have been processed, the best estimate of R can be obtained from this histogram. Bagnuolo states that modest amounts of photon and detector noise can be accommodated by this technique. He also reports computer simulations in which this algorithm attains a SNR approximately ten times greater than that of other astronomical speckle

imaging techniques. Application of the Fork algorithm to speckle images of Capella has been reported (Bagnuolo, Jr. and Sowell 1988).

de Freitas *et al.* (1987) (*cf.* de Freitas and Dainty 1988a) introduce an algorithm for reconstructing a randomly translating object viewed under conditions of severe photon-limiting. Since their algorithm has similarities to SAA it is mentioned here. They assert that, when the average number photons per speckle image, $\bar{N}_{\mathrm{p}}$, is less than ten, the SAA algorithm reference pixel cannot be located reliably. However, if it is known that a single, randomly translating, object is being viewed, the centroid of the photons provides a useful estimate of the shift position. By taking many photon-limited versions of the object, and SAAing them with the centroid as the shift position, a reconstruction of the object can be generated.

de Freitas *et al.* (1987) show that is useful to 'bin' the photon-limited images, by placing all the images with the same number of photons (N_{p}) in the same bin. SAA is then separately applied to the ensemble of images contained in each bin to produce several binned SAA images. de Freitas *et al.* (1987) note that recursive relations link the phase differences in the object's spectrum to quantities that can be derived from the binned SAA images. These relations can be exploited to recover an estimate of the phase of the object's spectrum from the SAA images. The resulting phase can be combined with the magnitude of the spectrum, obtained, for instance, from the SAA image for the bin $N_{\mathrm{p}} = 2$. An estimate of the spectrum results, which can be inverse Fourier transformed to give a reconstruction of the object. de Freitas *et al.* (1987) apply their algorithm to 80,084 computer-simulated one-dimensional photon-limited ($\bar{N}_{\mathrm{p}} = 3$) images of a randomly translating binary star. The separation and brightness ratio of the component stars of the binary were faithfully reconstructed by the algorithm.

de Freitas and Dainty (1988b) show that the centroiding technique can be considered to be a phase closure technique. Consequently, centroiding can, in principle, be generalized to allow diffraction-limited information about the phase of the object to be obtained, even when the object is viewed through atmospheric turbulence. They note, however, that the technique has a low theoretical SNR so one would expect it to be severely affected by noise.

The above processing is reminiscent of the 'nearest-centroid' shifting strategy considered by Sinton (1986, §5.4.2) when applying SAA to photon-limited speckle images. In this strategy the brightest pixel nearest to the centroid of all the brightest pixels in a speckle image is selected as the SAA reference pixel. He finds that, when SAA is applied to severely photon-limited speckle images of a binary star, SAA resolves the components of the star although the brightness ratio is usually significantly in error.

4.8.4 Non-astronomical Applications of Shift-and-add

Although somewhat out of place in a discussion of astronomical imaging techniques, the application of SAA to fields outside astronomy falls within the deconvolution theme of this thesis. Hence, two fields in which SAA has shown potential are briefly mentioned here.

Ultrasonic images generated by, first irradiating an object with ultrasound and, second, appropriately processing measurements of the resulting scattering, show a speckly nature similar to laser speckle (Abbott and Thurstone 1979; Gehlbach and Sommer 1987). However, in contrast to the astronomical speckle imaging situation, the distortion in ultrasonic images does not usually change rapidly and randomly with time. Temporal changes can be observed over long periods but the distortion

can typically be considered temporally invariant. Therefore, it is not possible to obtain an ensemble of independently blurred images by taking short exposure images separated in time. It transpires, however, that the distortion tends to be a strong function of frequency (Abbott and Thurstone 1979, §VI; Bates and Minard 1984). Thus, by gathering data in different, narrow frequency bands, an ensemble of approximately isoplanatically blurred images results, with the frequency-dependent psf being different for each image.

Radiation used in ultrasonic imaging is usually spatially coherent. This implies that it is possible to make direct measurements of the phase. Since the frequency-dependent psf has essentially random phase, the brightest version of the object in each speckle image has a random phase shift. Hence, the averaging process, inherent in basic SAA, results in destructive interference obscuring the object. This can be overcome by shifting the phase of every pixel in the image such that the phase at the image's brightest pixel becomes zero. Then, when SAA is applied to the phase-shifted images, the brightest versions of the object reinforce. This extension to basic SAA is called *coherent SAA* (Bates and Robinson 1982; Robinson 1982; Minard *et al.* 1985). The coherent SAA image $f_{\mathrm{csa},M}(\mathbf{x})$ is described by (Sinton 1986)

$$f_{\mathrm{csa},M}(\mathbf{x}) = \langle s_m(\mathbf{x} - \mathbf{x}_m)\mathrm{e}^{\imath\phi(\mathbf{x}_m)}\rangle_M, \tag{4.55}$$

where $\mathbf{x}_m$ is again the position of the pixel with greatest magnitude, and $\phi(\mathbf{x}_m) = \mathcal{P}[s_m(\mathbf{x}_m)]$. The other versions of the object and the contamination are also coherent and therefore have a random phase from speckle image to speckle image. Thus, these effects average toward zero instead of towards a positive fog (§4.8.1), so procedures for fog removal are usually not required.

Another non-astronomical field in which SAA has been invoked is speech processing. Much speech processing research is based on the 'source-filter' model of speech production, in which the larynx generates an excitation function that is filtered by the vocal tract to produce the spoken sound (Markel and Gray, Jr. 1976, §1.3).

Speech production can be conveniently classified into two broad classes depending upon the function of the glottis. In *voiced speech*, the excitation consists of a sequence of quasi-periodic pressure pulses, called *glottal pulses*, induced by the glottis. Such speech can be represented by the convolutional model

$$s(t) = g(t) \odot v(t) + c(t), \tag{4.56}$$

where $s(t)$ is the spoken speech, $g(t)$ represents the glottal pulses, $v(t)$ is the *vocal tract response* and $c(t)$ incorporates all departures of the convolutional model from reality. For *unvoiced speech* the glottis no longer produces quasi-periodic glottal pulses and the excitation function is essentially random air turbulence.

Since voiced speech is approximately repetitive at the pitch frequency (this is apparent in Fig. 4.12(a)), it can conveniently be broken into segments, each of about one pitch period in length. Thus,

$$s(t) = \sum_{m=1}^{M} s_m(t - T_m), \tag{4.57}$$

where T_m is the time at which the m^{th} speech segment $s_m(t)$ occurs. Each such segment can then be represented by

$$s_m(t) = g_m(t) \odot v_m(t). \tag{4.58}$$

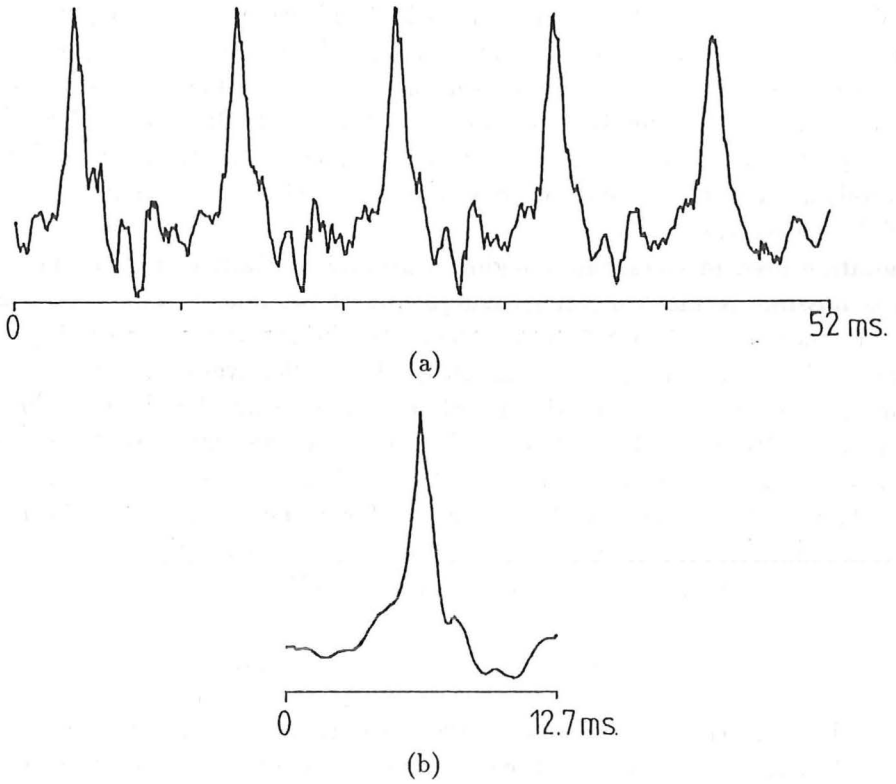


Figure 4.12: SAA processing of voiced speech to extract the average glottal pulse: (a) voiced speech extending over several pitch periods; (b) $\hat{g}(t) = f_{\text{sa}}(t)$ resulting from applying SAA processing to 300 pitch periods of speech data.

The vocal tract varies as different sounds are spoken, while the glottal pulse is relatively constant for a given speaker. Brieseman *et al.* (1987) note that if one could validly assume that $g(t)$ was temporally invariant, an estimate of the glottal pulse could be recovered by performing shift-and-add on the speech segments defined by (4.58) (see Fig. 4.12(b)). However, since the glottal pulse does change as different sounds are spoken (Miller 1959), and especially as the pitch is altered, shift-and-add does not extract the actual individual pulse shape. Rather, it produces an 'average' glottal pulse $\hat{g}(t)$ which is that part of the speech having a constant shape in each pitch period. Potential applications for this average glottal pulse include speech therapy, diagnosis of vocal tract disorders, speaker identification and speech recognition applications (Brieseman *et al.* 1987 *cf.* Bates *et al.* 1987b).

4.9 Speckle Imaging at Infrared

Many objects of interest to astronomers and astrophysists radiate most strongly in the infrared and near-infrared bands of electromagnetic spectrum (Roy and Clarke 1977, §14.8). It follows, therefore, that useful astrophysical information is often obtained by observing these objects at infrared wavelengths. Speckle imaging techniques have been extended to the infrared (Sibille *et al.* 1979; Howell *et al.* 1981; Mariotti *et al.* 1983; Dyck and Howell 1985), to produce an effective increase in the resolution attainable by earth-based telescopes operating at these wavelengths.

Atmospheric characteristics are wavelength-dependent (Roddier 1981, p. 301). For example, experiments conducted by (Mariotti 1983) reveal that, at $\lambda = 4.6\mu$m,

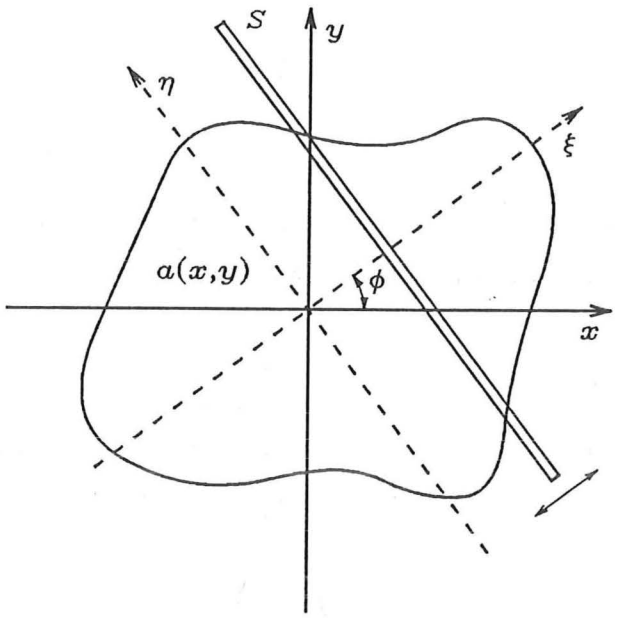


Figure 4.13: One-dimensional infrared speckle image capture. The two-dimensional speckle image $a(x, y)$ is scanned across long, narrow slit (S), in a direction (ξ) perpendicular to the length of the slit. All the energy passing through the slit is integrated by a single infrared detector whose time varying output provides the one-dimensional speckle image $a(\xi)$.

r_0 has a typical value of 1.2m, compared with 10cm at visible wavelengths (§4.2). A comparison of the critical parameters describing the formation and capture of speckle images at the visible wavelength of 500nm and the infrared wavelength of 5μm are presented in Table 4.1. Inspection of the second column of this table shows that r_0, τ_r, and the isoplanatic angle all scale with wavelength as $\lambda^{6/5}$ whilst $\Delta\lambda$ scales as λ^2 (*cf.* Roddier 1988, Table 1).

Technical difficulties in constructing electronic cameras to operate at infrared wavelengths mean that techniques for recording infrared speckle images have tended to differ from the corresponding techniques for visible wavelengths. At visible wavelengths a two-dimensional array of detectors is typically employed, while at infrared wavelengths only a single detector has been available until quite recently. Considerable research into detector arrays to operate at infrared wavelengths has, however, been reported (*cf.* Monin *et al.* 1987) and several prototype devices have already been tested (M^{c}Lean 1988) making direct two-dimensional infrared speckle imaging possible.

Discussion of the capture of infrared speckle images in this thesis is restricted to an implementation that has proven valuable over the last decade or so. This implementation employs a single infrared detector, and is now discussed with reference to Fig. 4.13. Since only a single infrared detector is available for detecting the infrared radiation, speckle images are typically recorded by rapidly scanning the two-dimensional speckle image across a long narrow slit positioned in the focal plane of the telescope, in front of the detector (Sibille *et al.* 1979). At each instant, the radiation passing through the slit is focussed upon, and thus integrated by, the detector. As the image is scanned in the ξ direction, the time-varying output of the detector generates a one-dimensional speckle image as a function of ξ. When the slit has passed over the entire image the complete one-dimensional speckle image is

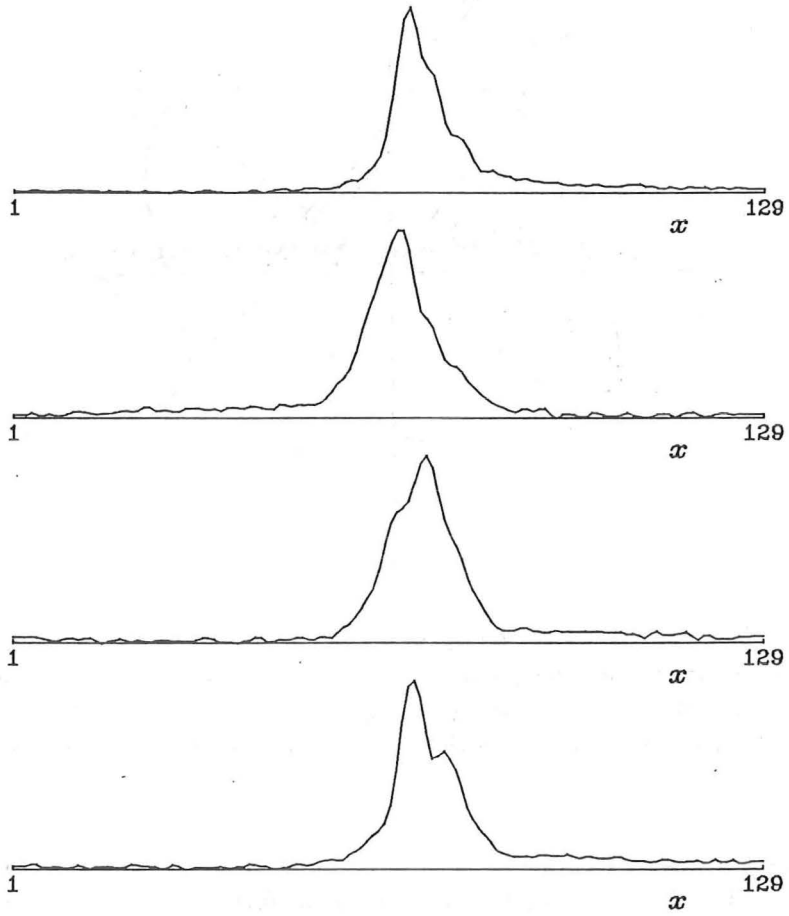


Figure 4.14: Four infrared speckle images typifying those obtained for the unresolved star SAO 133312. ($\lambda_0 = 2.2\mu$m, $\Delta\lambda = 0.6\mu$m, $D = 2.28$m, scan duration = 76ms). Images courtesy of Steward Observatory.

generated. This image is sometimes called a *speckle scan.* One-dimensional infrared speckle images of the unresolved star SAO 133312 are shown in Fig. 4.14.

The theory of image reconstruction from projections (see §2.4.1) ensures that if a two-dimensional image can be expressed as a convolution, so too can a one-dimensional projection of that image. Thus, the one-dimensional infrared speckle image, which is the projection of the two-dimensional image, is also expressible in the form of (4.4). This is also discussed in §2.4.1.

The principle requirements for accurately recording the one-dimensional speckle images are (Chelli *et al.* 1979)

(i) the slit should have a width $W = 1/(2h)$ where h is the highest spatial frequency to be measured,

(ii) the slit must be sufficiently long that the entire two-dimensional speckle image is integrated, and

(iii) the image must be scanned across the slit in a time less than τ_s (typically 50–100arcsec/s).

The scanning motion is typically generated by linear tilts of the Cassegrain an-

tenna's secondary mirror. To preserve the spatial frequency content of the two-dimensional speckle image, the slit must have an angular width less than the diffraction limit of the telescope and the secondary mirror scan motion must be highly linear. Unfortunately, non-linearities in the scanning mechanisms of currently available speckle recording systems introduce significant contamination (Leinert and Dyck 1983; McCarthy *et al.* 1987). Methods for compensating for the nonlinearity have been proposed (*cf.* McCarthy *et al.* 1987) and a new technique (Davey *et al.* 1989b), capable of compensating for this non-linear scanning motion, is introduced in §6.3.

Several of the optical speckle processing techniques discussed in §§4.5–4.8 have been applied to one-dimensional infrared speckle images. For example, speckle interferometry (Sibille *et al.* 1979; Dyck and Howell 1985) and triple correlation (Christou *et al.* 1987b) (*cf.* Freeman *et al.* 1988) has been applied to infrared speckle images of several astronomical sources. McCarthy and Cobb (1986) discuss deconvolution techniques for infrared speckle imaging. Freeman *et al.* (1987) compare Knox-Thompson, triple correlation, and shift-and-add techniques as they apply to infrared speckle images.

Chapter 5

Fourier Phase and Magnitude Retrieval Algorithms

In a great variety of branches of engineering science it is either impossible or impractical to make direct measurements of the quantity in which one is interested. One is forced to infer the quantity from data obtained by measuring another quantity, related to the desired quantity in some particular way. A relation that frequently arises is Fourier transformation, with the desired and directly measured quantities here called, respectively, the image and its spectrum. Table 5.1 lists the analogues of image and spectrum for a number of relevant technical sciences.

If all the pertinent characteristics of the directly measured quantity are accurately recordable, the desired quantity can be immediately reconstructed. There are, however, many important and commonly occurring situations in which it can be inconvenient or even infeasible to record certain highly significant features of the measured quantities. For instance, the phases of spectra are usually much more difficult to measure accurately than their intensities. In several of the technical sciences listed in Table 5.1, for instance, it is almost impossible to make usefully accurate direct recordings of phases.

In recent years it has become apparent that the Fourier-space phase and magnitude are uniquely linked (Bruck and Sodin 1979; Hayes 1982; Lane 1988). For instance, if the spectral magnitude (i.e. the magnitude of its spectrum) of an image is discarded, and replaced with a unity magnitude, features of the original image are preserved (Oppenheim and Lim 1981). This phase dominance is discussed, and illustrated, in §5.1.2. It follows from this unique linkage between the spectral magnitude and phase, that it is possible, at least in theory, to completely infer the quantity of interest from the limited directly-recorded spectral information. The proviso here, of course, is that the limited information is recorded with sufficient accuracy to allow this.

Technical science	Image	Spectrum
Antenna engineering	Aperture distribution	Radiation pattern
X-ray and neutron crystallography	Molecular structure	Diffraction pattern
Communication engineering	Signal	Frequency spectrum

Table 5.1: Typical terminology applied in several technical sciences to describe an image or signal, and its Fourier transform. In several of these sciences it is significantly easier to measure spectra than images. (Refer also to Table 2.2.)

Consider, for example, a situation in which only measurements of the spectrum of the quantity of interest can be made. Furthermore, in this hypothetical situation, the magnitude of the spectrum is accurately recordable, but no measurement of the phase can be made. The image generated from only knowledge of the spectral magnitude, and assuming the spectral phase is zero, is severely degraded. Thus, in order to accurately recover the image, one must know both the spectral magnitude and phase. An algorithm that would exploit the unique relationship between the magnitude and phase, thereby recovering the unknown phase, would allow an accurate reconstruction to be generated. Other situations arise in which one measures, or otherwise infers, the spectral phase when no estimate of the magnitude is available (Hayes 1982; Lane and Bates 1987b). In such a situation it would be useful to recover the magnitude from the known phase. It is algorithms for implementing the recovery of the phase (magnitude), when only the magnitude (phase) is given, that are the subject of this chapter.

The motivation for the development of algorithms capable of recovering the spectral phase or magnitude, and the reason for reviewing them in this thesis, are discussed in §5.1.

A number of practical iterative algorithms for recovering the spectral phase from the Fourier magnitude have been proposed (Fienup 1978; Fienup 1982; Fright 1984; Lane 1988). Several of the iterative algorithms that have proved effective in practice are noted in §5.2. A few of these algorithms are discussed in greater depth in §§5.2.1 and 5.2.2. This is appropriate since these chosen algorithms provide a useful introduction to several of the general blind deconvolution algorithms discussed in Chapter 7.

Direct (i.e. non-iterative) approaches to recovering the spectral phase have also been investigated (*cf.* Lane 1988), and are discussed in §5.3. Unfortunately these approaches tend to be highly susceptible to contamination and therefore, as yet, have limited usefulness in practical situations. They have, however, added valuably to the debate over the uniqueness of Fourier phase retrieval in more-than-one dimensions.

Iterative algorithms to implement the recovery of spectral magnitude from spectral phase are discussed in §5.4. This problem has traditionally been considered the easier of the two problems (Hayes 1982) because the spectral phase is known to dominate the magnitude, at least as regards the form of a reconstructed image (Oppenheim and Lim 1981).

The majority of the material presented in this chapter reviews established algorithms for phase and magnitude retrieval, although the computational examples of various algorithms are original, as is the detailed study of the affect of contamination upon the zero-sheet of an autocorrelation (§5.3.4). This pictorial study provides considerable insight into the contamination induced distortion of zero-sheets in the vicinity of an intersection in four-dimensional space.

5.1 Motivation

5.1.1 Motivation for Fourier Phase Retrieval

There are a number of reasons why, in practical image recording situations, the phase may be difficult to measure accurately. Following Lane (1988, §2.4) these can be summarized as:

- Accurate receivers sensitive to high frequency phase information may be expensive or difficult to obtain (Morris 1985).

- The phase information may be distorted during propagation. This is the case with atmospheric distortion (Chapter 4).
- The instrument may have inherent inaccuracies which cause a distortion in the measured phase, e.g. a feed defocus or misaligned panels of an antenna (Morris 1985).
- The phase may be discarded to increase storage and/or transmission efficiency. This occurs in LPC coding of speech signals (Makhoul 1975).
- There may be difficulties in maintaining stable phase references, as occurs in Very Long Baseline interferometry where measurements are made at widely spaced geographic locations (Readhead *et al.* 1980).

The deterioration of, or loss of, the phase information due to one, or a combination of, the above factors may have undesirable consequences which it would be beneficial to remove.

In recent years it has been demonstrated that a unique relationship almost always exists between the spectral magnitude and the phase in more-than-one-dimensional space (Lane 1988, §4.3). This question of uniqueness is addressed in pioneering work by Bruck and Sodin (1979) and Hayes (1982), who argue that such a relationship should exist for a discrete compact image. The existence of such a unique relationship linking the spectral magnitude and phase leads to the conclusion that it is theoretically possible to recover the phase from the magnitude (Bates 1982a; Lane 1988). This problem is now often known as the *Fourier phase problem* (Fright 1984) and is posed as:

> "Given only $|Q(\mathbf{u})|$ (or, equivalently, $|Q(\mathbf{u})|^2$), recover $\mathcal{P}[F(\mathbf{u})]$ (or, alternatively, $q(\mathbf{x})$)."

It is useful (Fright 1984) to divide the phase problem into two further categories, the *pure phase problem* in which only the spectral magnitude is available, and the *partial phase problem* where additional *a priori* information, e.g. some estimate of the spectral phase, is provided. Although arising less often in practice, the main emphasis of this chapter is on the pure phase problem, since computational solutions to partial phase problems are usually constructed easily enough by appropriate adaptation of procedures developed for the pure phase problem.

One further requirement for Fourier phase retrieval is that sufficient Fourier-space data are collected (Bates and M^{c}Donnell 1986, §20). The necessary condition is that the spectral magnitude must be sampled at a rate greater than or equal to twice the Nyquist sampling rate for the object (Bates and Mnyama 1986, §II.C), i.e. the Fourier magnitude is *oversampled* by a factor of two. There are some phase problems of Fourier type for which the oversampling condition is not met. One example is the crystallographic phase problem in which the molecular structure of a crystal is to be inferred from an observed x-ray diffraction pattern (Fright 1984, §2.4; Bates and Mnyama 1986, §II.E). The inverse Fourier transform of the intensity of such a spectrum yields an aliased form of the autocorrelation, known as the *Patterson* (Ramachandran and Srinivasan 1970). Crystallographic spectra are discrete, existing only for the Nyquist samples, so it is not possible to sample these spectra at twice the Nyquist rate. Since the oversampling criteria for Fourier phase retrieval is not satisfied, the crystallographic phase problem is not a Fourier phase problem, in the sense that that term is used in this thesis.

Further details of some physical recording situations involving the phase problem are listed in Table 5.2.

Application	Causes	Type
X-ray and neutron crystallography	Wavelength to short for sensing	Pure but non-Fourier
Electron microscopy	Instrument imperfections Wavelength too short for sensing	Partial
Astronomical speckle imaging	Turbulent medium	Partial
Astronomical speckle interferometry	Turbulent medium	Pure
Radio engineering	Incorrect measurement (wavelength dependent)	Partial
Ultrasonics Acoustics	Distorting medium	Partial
Communications Speech processing	Economy of storage	Pure or Partial

Table 5.2: Typical technical sciences involving the phase problem.

In one dimension, no unique relationship exists between the spectral magnitude and phase (Fright 1984, §3.3). Although it can on occasion be useful to attempt one-dimensional phase retrieval, the one-dimensional situation is not reviewed here. Hence, in the following discussion it is assumed that $K > 1$ and that $\mathbf{x}$ and $\mathbf{u}$ span two-or-more-dimensional spaces. The one-dimensional phase problem has, however, been discussed in depth from a number of viewpoints (Walther 1963; Bates 1969; Burge *et al.* 1976; Fright 1984).

It is important to recognize, when considering Fourier phase retrieval algorithms, that the spectra of images having the same image-form (§2.1.5) all have the same magnitude. Thus, it follows that, when presented with only the Fourier magnitude, one can, at best, only recover the image-form of the object. This fundamental ambiguity is usually of little account, however, because an image is essentially unchanged either by rotating it through 180° (and conjugating its phase) or by altering its position. Consequently, it is assumed in the following sections that it is the object's image-form which is reconstructed by Fourier phase retrieval algorithms.

Algorithms for implementing phase retrieval can be split into two broad classes, those that only work for special types of objects, and those that work for general objects. It is algorithms capable of reconstructing general objects that are of principle interest to this thesis, although several of the algorithms that exploit special features of particular classes of objects are mentioned in passing here.

Consider an object that includes an unresolvable part that is separated from the rest of the object by a distance that is greater than the extent (§2.1.4) of the rest of the object. The autocorrelation theorem (§2.4) states that by inverse Fourier transforming the intensity of the spectrum one obtains the autocorrelation of the object. It follows that, because of the isolated reference pixel in the object, the object can be

found as an isolated part of the autocorrelation (Napier and Bates 1971; Bates and McDonnell 1986, §21), thereby implementing phase retrieval. Other algorithms for objects with two or more isolated unresolved parts, which can, in general, be arbitrarily close to the remainder of the image, have also been successfully implemented (Fiddy *et al.* 1983; Fienup 1983). Still other algorithms for implementing Fourier phase retrieval have been realized for objects having certain specific support shapes, which must be known exactly (Arsenault and Chalasinska-Macukow 1983; Crimmins and Fienup 1983; Brames 1986). Techniques for deriving the exact support of a positive object from its spectral magnitude have been reported (Brames 1987).

The motivation for reviewing Fourier phase retrieval algorithms within this thesis is twofold. Firstly, phase retrieval is a special case of blind deconvolution. This becomes apparent if one squares the given $|Q(\mathbf{u})|$ and invokes the autocorrelation property of the Fourier transform (see Table 2.3) to obtain

$$|Q(\mathbf{u})|^2 \longleftrightarrow \mathcal{A}[q(\mathbf{x})] = q(\mathbf{x}) \circledcirc q^*(-\mathbf{x}) = q^*(-\mathbf{x}) \circledcirc q(\mathbf{x}). \tag{5.1}$$

Comparing this equation with (1.4), or (3.1), indicates that $q^*(-\mathbf{x})$ plays the part of the psf in (5.1). Since the form of $q(\mathbf{x})$ is initially unknown, then so is the form of $q^*(-\mathbf{x})$, because both quantities, of course, possess the same image-form. Consequently, it follows that the recovery of $q(\mathbf{x})$ from $|Q(\mathbf{u})|$ is a blind deconvolution problem (see §1.5). Secondly, the general blind deconvolution algorithms discussed in Chapter 7 can be considered to be extensions of the Fourier phase retrieval methods outlined in this chapter. Thus, discussion of phase retrieval algorithms presented in §§5.2 and 5.3 prepares the reader for the general blind deconvolution algorithms presented in Chapter 7.

5.1.2 Motivation for Fourier Magnitude Retrieval

Attention is now turned to the motivation for pursuing algorithms for the recovery of magnitude from phase, which is a complementary problem to the Fourier phase problem. Algorithms for overcoming this, often called the *Fourier magnitude problem*, have received intensive theoretical study (*cf.* Oppenheim and Lim 1981; Hayes 1982; Bruck and Sodin 1983). Practical algorithms implementing the recovery of magnitude from phase have also been reported (Oppenheim *et al.* 1982; Lane and Bates 1987b; Lane 1988).

The dominance of the phase over the magnitude, as regards the form of an image, has been noted by many authors (e.g. Ramachandran and Srinivasan 1970; Oppenheim and Lim 1981; Hayes 1982; Fright 1984; Bates and Mnyama 1986; Lane 1988). This *phase dominance* is apparent in the example contained in Fig. 5.1. When the true spectral phase is combined with a magnitude distribution containing no information about the original image (it is chosen to be unity in fact), characteristics of the original image are still apparent in the 'phase-only' reconstruction (see Fig. 5.1(b)). However, when the true spectral magnitude is combined with a random (Fig. 5.1(c)) or zero (Fig. 5.1(d)) phase, the 'magnitude-only' reconstruction does not contain any features evident in the original image. Oppenheim and Lim (1981) show that a reconstruction obtained by combining the true phase with an approximate estimate of the magnitude is significantly more faithful than the reconstruction generated from the true phase combined with a unity magnitude.

Comparison of Fig. 5.1(a) and (b) reveals that the phase-only reconstruction emphasises the edges of the original image. This edge enhancement is characteristic of high-pass filtering (Gonzalez and Wintz 1977, §4.4.2). Lane (1988) notes that the spectral magnitude of a typical image is usually low-pass in nature (see also §3.3).

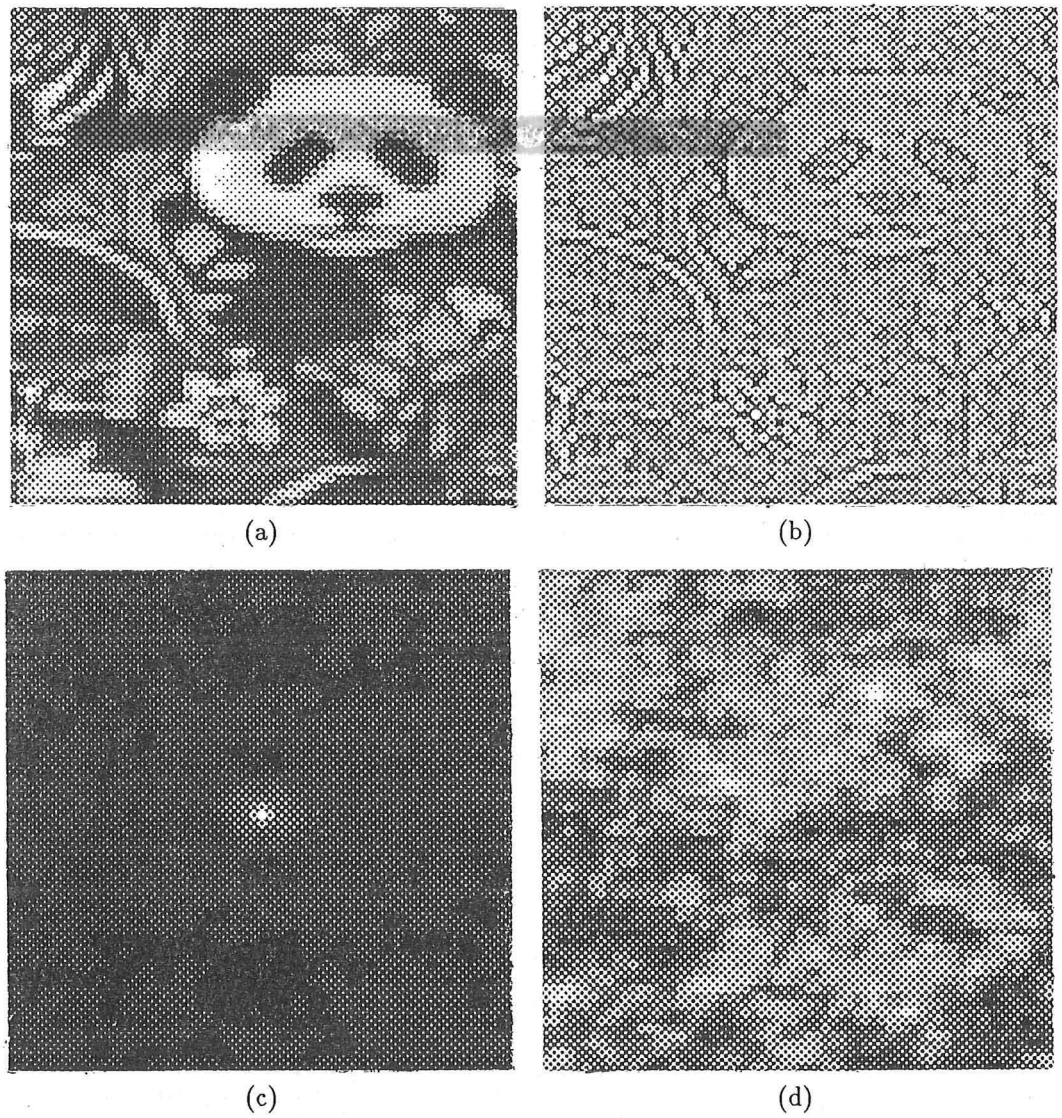


Figure 5.1: Dominance of the phase of an image in determining the image-form: (a) the original image $q(\mathbf{x})$; (b) the image resulting from combining $\mathcal{P}\left[Q(\mathbf{x})\right]$ with a unity magnitude; (c) the image resulting from combining $|Q(\mathbf{x})|$ with a zero phase; (d) the image resulting from combining $|Q(\mathbf{x})|$ with a random phase. Note that considerably more detail apparent in the original image is apparent in (b) than in (c) or (d). Lane (1988, §2.5) notes that, when displaying the magnitude of a function that has had its dc value removed, it is necessary to add an offset to the magnitude before it is displayed. If this precaution is not taken, discontinuities arise at the zero-crossings of the magnitude of the bipolar image. Such an offset was added to (b) before display.

Therefore, replacing it with unity magnitude is effectively enhancing the spectral magnitude at high frequencies. This is, by definition, the function of a high pass filter, so it is not surprising that the edges of the original image are enhanced in Fig. 5.1(b).

The Fourier magnitude problem arises in several practical situations. For example, consider the situation in which it is required to solve the blind deconvolution problem posed in §1.5, for the special case in which the spectral phase of the psf is known to be zero. Since the blurred image is the convolution of the true object and a psf, it follows from the convolution theorem (Table 2.3) that the spectra of the object and the psf are multiplied. Furthermore, since, in this special case, the spectral phase of the psf is zero, the phase of the blurred spectrum must be the object's spectral phase. The spectral magnitude, however, can not be reconstructed directly from the measured data, so the object is not directly computable. However, a magnitude retrieval algorithm applied to the spectral phase would recover the object, thereby implementing blind deconvolution. This application of magnitude retrieval algorithms is outlined in §5.4 and illustrated in §7.2. The magnitude problem also arises in other areas, for example, phase-only holograms, or kinoforms (Lesem *et al.* 1969), and Fourier synthesis of crystal structure (Ramachandran and Srinivasan 1970).

Although not in itself a deconvolution problem, the Fourier magnitude problem is encountered in one of the blind deconvolution algorithms discussed in Chapter 7. Consequently algorithms implementing the recovery of the magnitude from the known phase are introduced in §5.4.

5.2 Iterative Phase Retrieval Algorithms

Fourier phase retrieval has received a considerable amount of research effort in recent years, with many algorithms being realized. The algorithms that have proven most effective in practice are iterative. Several of these algorithms are discussed in §§5.2.1 and 5.2.2.

The iterative phase retrieval algorithms that I have chosen to discuss in more depth are based upon the *general iterative processing loop* (Fienup 1982; Bates and Davey 1988; Lane 1988, §5.1) depicted in Fig. 5.2. Limiting review of iterative phase retrieval algorithms to a few that are based on this loop is appropriate for several reasons. Firstly, these algorithms have received widespread study and are perhaps the most widely applied of all phase retrieval algorithms. Secondly, the iterative blind deconvolution algorithms discussed in §§7.3 and 7.4 employ an extended form of the general processing loop, so that material presented here constitutes a useful introduction to those techniques. Lastly, a detailed review of phase retrieval algorithms is outside the scope of this thesis since phase retrieval is but a small aspect of blind deconvolution. Several recent reviews of the status of Fourier phase retrieval are available (Fienup 1984; Fright 1984; Dainty 1984; Bates and Mnyama 1986; Lane 1988; M^{c}Callum 1989).

It is worth mentioning, in passing, that the loop depicted in Fig. 5.2 has also been usefully applied to retrieve information in fields other than phase retrieval. For instance, it has found application in the magnitude problem (Hayes 1982) (see also §5.4) and bandlimited extrapolation (Gerchberg 1974; Sanz and Huang 1984; Kani and Dainty 1988).

The essence of the general iterative processing loop is to transform an estimate of the object back and forth between Fourier-space and image-space applying known

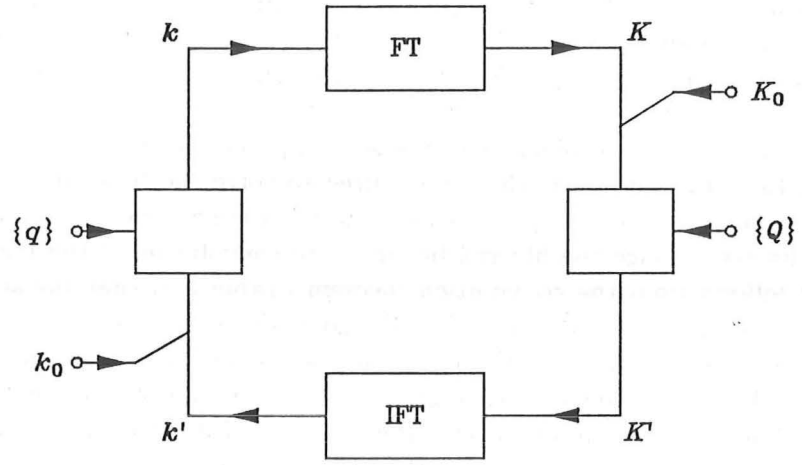


Figure 5.2: The general loop that forms the basis of several iterative Fourier phase and magnitude retrieval algorithms.

constraints in each space at each iteration. The role of the constraints is to urge the reconstruction at each step to become more faithful. In this loop the constraints at each step comprise incomplete information about either the image $q(\mathbf{x})$ or its spectrum $Q(\mathbf{u})$. In this thesis these image-space and Fourier-space constraints are denoted by $\{q\}$ and $\{Q\}$, respectively. The i^{th} estimate of the object and its spectrum are denoted by $k_i(\mathbf{x})$ and $K_i(\mathbf{u})$ respectively. In addition a set of starting data, or initial conditions, are provided, either in image-space as $k_0 = k_0(\mathbf{x})$ or in Fourier-space as $K_0 = K_0(\mathbf{u})$. The goal of the iterations is to progressively urge the initial estimate of the object to become more faithful.

The Fourier-space constraint, in the case of the Fourier-phase problem, is the known magnitude of the spectrum, i.e. $|Q(u,v)|$. In addition, if any partial information about $\mathcal{P}[Q(u,v)]$ is known, it can also be incorporated into $\{Q\}$. Image-space constraints that are commonly applied in conjunction with the general iterative processing loop are:

(i) A *support constraint* when it is known that the image is compact and an estimate of its support is known. Any pixel of non-zero value lying outside the estimate of the support violates this constraint. Unless otherwise specified, in this thesis the estimate of the support is taken to be the image-box (§2.1.4) of the image (i.e. a box, with sides parallel to the chosen coordinate axes, that just encloses the image).

(ii) A *reality constraint* when it is known that the image is real. Any pixel whose value has a non-zero imaginary part violates this constraint.

(iii) A *positivity constraint* when it is known that the image is positive. Any pixel that violates the reality constraint also violates this constraint. In addition the real part of the each pixel must be non-negative to satisfy this constraint.

These image-space constraints can either be applied separately, or in combination.

Various other iterative Fourier phase retrieval methods, which are not based on the loop depicted in Fig. 5.2 have been proposed. Two of these methods are now mentioned in passing.

Maximum entropy (Jaynes 1982; Nityananda and Narayan 1982) image restoration has been applied with considerable success to the phase problem (Gull and

Daniell 1978; Bryan and Skilling 1986; Narayan 1987). Maximum entropy algorithms constrain the estimate of the object by attempting to maximize the entropy. Thus, in general terms, the maximum entropy algorithm tends to reconstruct the smoothest object consistent with the given data. It is a matter of some debate as to whether the entropy has the form

$$E_0 = \int \ln[f(\mathbf{x})]\,\mathrm{d}\mathbf{x}, \tag{5.2}$$

or

$$E_0 = \int f(\mathbf{x}) \ln[f(\mathbf{x})]\,\mathrm{d}\mathbf{x}. \tag{5.3}$$

Nityananda and Narayan (1982) show that maximum entropy image restoration produces the most faithful restorations when objects consist of isolated peaks on a flat background.

Nieto-Vesperinas *et al.* (1988) (*cf.* Nieto-Vesperinas and Mendez 1986) outline a phase retrieval algorithm based on the simulated annealing algorithm of Kirkpatrick *et al.* (1983). Simulated annealing is a Monte Carlo method for solving large systems of equations, which has developed from a procedure devised by Metropolis *et al.* (1953). Nieto-Vesperinas *et al.* (1988) report that the algorithm is robust and flexible and reconstructs high contrast areas faithfully, but unfortunately the algorithm is very computationally expensive. They suggest that the technique could usefully complement other iterative algorithms, for example those of (§5.2.2), in situations where those algorithms stagnate.

5.2.1 The Gerchberg-Saxton Algorithm

Gerchberg and Saxton (1972) put forward an iterative algorithm for solving the Fourier phase problem. This, the Gerchberg-Saxton (GS) algorithm, is based on the *general iterative processing loop* depicted in Fig. 5.2. For their technique both the object and its spectrum can, in general, be complex. In terms of the notation of Fig. 5.2, the image-space constraint in the GS algorithm is the known magnitude of the image (i.e. $\{q\} = |q(\mathbf{x})|$), while the Fourier-space constraint is the measured Fourier magnitude (i.e. $\{Q\} = |Q(\mathbf{u})|$). Thus, the GS algorithm is a partial phase problem since the image-space magnitude, in addition to the spectral magnitude, must be known.

Iterations commence with an estimate of the image-space phase $k_0(\mathbf{x})$, which is often conveniently chosen to be pseudo-randomly distributed between $-\pi$ and π (Gerchberg and Saxton 1972). The first estimate of the object $k_0(\mathbf{x})$ is given by

$$k_1(\mathbf{x}) = |q(\mathbf{x})| e^{i\mathcal{P}[k_0(\mathbf{x})]}. \tag{5.4}$$

One iteration of the GS algorithm is described by

$$K_i(\mathbf{u}) = \mathcal{F}[k_i(\mathbf{x})], \tag{5.5}$$

$$K_i'(\mathbf{u}) = |Q(\mathbf{u})| e^{i\mathcal{P}[K_i(\mathbf{u})]}, \tag{5.6}$$

$$k_i'(\mathbf{x}) = \mathcal{F}^{-1}[K_i'(\mathbf{u})], \tag{5.7}$$

$$k_{i+1}(\mathbf{x}) = |q(\mathbf{x})| e^{i\mathcal{P}[k_i'(\mathbf{x})]}. \tag{5.8}$$

An iterative loop is formed by taking the most recent $k(\mathbf{x})$ to be the new $k(\mathbf{x})$ (i.e. the LHS of (5.8) becomes the RHS of (5.5)). Iterations proceed until either convergence is obtained, or the algorithm ceases to converge.

The GS algorithm can not diverge although convergence is not guaranteed (Gerchberg and Saxton 1972). In many situations the iterations converge very slowly or not at all. When this happens the algorithm is said to have reached *stagnation.* In practice stagnation is a serious impediment to effective phase retrieval. Gerchberg (1986) terms stagnation the *lock problem* and proposes a solution to it that is somewhat reminescent of the hybrid input-output technique proposed by Fienup (1982), which is discussed in §5.2.2.

The GS algorithm was originally intended to be applied in electron microscopy contexts, although it has also found application in other fields. For example, (Gardenier *et al.* 1986; Bates *et al.* 1987a) report modifying the technique to detect deformations in radio antennas employed in satellite communication, thereby allowing the antennas to be aligned correctly. Deformation of an antenna's surface leads to phase perturbations in the antenna aperture distribution, and therefore to increased far-field sidelobe levels (Blake 1984). By calculating the antenna aperture magnitude from design data (to provide $\{q\}$), and measuring the far-field radiation pattern ($\{Q\}$), the GS algorithm can be invoked to estimate the severity and location of the deformations. The GS technique has also found application in acoustic microscopy (Bates *et al.* 1987a; Fright *et al.* 1988).

5.2.2 Fienup's Algorithms

Fienup (1978) has instigated a type of processing, which is essentially an adaptation of the GS algorithm, to solve the Fourier phase problem from a single intensity measurement. Several algorithms have subsequently been developed by Fienup (1982), each of which applies the first three steps of the GS algorithm (i.e. (5.5)–(5.7)). However, in Fienup's algorithms, the fourth step of the GS algorithm is replaced with a step that incorporates additional image-space constraints. Two of Fienup's algorithms are described here.

In the simplest of Fienup's algorithms, called *error-reduction*, the fourth step in the GS algorithm (i.e. (5.8)) is replaced by

$$k_{i+1}(\mathbf{x}) = \begin{cases} k'_i(\mathbf{x}) & \text{if } \mathbf{x} \in \Upsilon \\ 0 & \text{otherwise,} \end{cases} \tag{5.9}$$

where, for convenience, Υ denotes the set of points in image-space where the current estimate $k'_i(\mathbf{x})$ satisfies the known constraints (i.e. $\{q\}$) in this space. The operation is non-expansive (Tom *et al.* 1981) in the sense that $k(\mathbf{x})$ cannot be further from the true solution than $k'(\mathbf{x})$ (Youla and Webb 1982). Consequently it is impossible for the error-reduction algorithm to diverge. However, a common problem encountered when applying the error-reduction algorithm is a slow convergence rate, or no convergence at all.

Some means of evaluating the convergence of the iterative loop is necessary. Ideally one would compare the latest reconstruction with the true image, to obtain an estimate of the fidelity of the reconstruction. This error measure, here called the *true error* and denoted E_T, is conveniently defined at the i^{th} iteration as

$$E_T = \frac{\int_{(K)}\int |f(\mathbf{x}) - k_i(\mathbf{x})|^2 \, d\mathbf{x}}{\int_{(K)}\int |f(\mathbf{x})|^2 \, d\mathbf{x}}. \tag{5.10}$$

When the true image is unknown, as it always is in the real world, it is not possible to calculate E_T, so some alternative error measure is needed. One such measure is

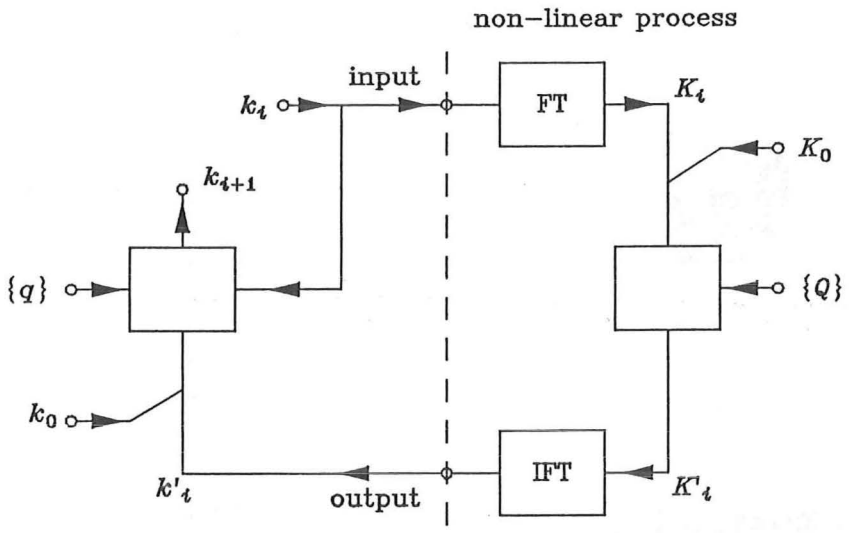


Figure 5.3: The hybrid input-output phase retrieval algorithm proposed by Fienup illustrating the 'input-output' concept. The section of the iterative loop to the right of the dotted line can usefully be considered to be a non-linear process.

given by the amount that the reconstruction violates the image-space constraints. This, the *image-space error* (E_I), is described by

$$E_I = \frac{\int(K)\int_{\mathbf{x}\notin\Upsilon} |k_i(\mathbf{x})|^2 \, \mathrm{dx}}{\int(K)\int |k_i(\mathbf{x})|^2 \, \mathrm{dx}}. \tag{5.11}$$

Alternatively, it is possible to determine how much the Fourier-space estimate violates the Fourier constraints. In the phase problem this error measure is defined by

$$E_F = \frac{\int(K)\int(|F(\mathbf{u})| - |K_i(\mathbf{u})|)^2 \, \mathrm{du}}{\int(K)\int |F(\mathbf{u})|^2 \, \mathrm{du}}. \tag{5.12}$$

It is the image-space error measure, E_I, that is calculated to assess the convergence of the iterative loop in examples of phase retrieval algorithms presented in this thesis.

The often slow convergence of the error-reduction algorithm can be accelerated by another of Fienup's iterative schemes, the *hybrid input-output* algorithm (Fienup 1982). This algorithm has the fastest rate of convergence of all the iterative schemes proposed by Fienup (Fienup 1982; Fright 1984, §4.2.2), and has proven to be effective in practice (Fienup 1982; Lane 1988). The algorithm is defined by (5.5)–(5.7) with the last step of the GS algorithm, i.e. (5.8), replaced by

$$k_{i+1}(\mathbf{x}) = \begin{cases} k'_i(\mathbf{x}) & \text{if } \mathbf{x} \in \Upsilon \\ k_i(\mathbf{x}) - \beta k'_i(\mathbf{x}) & \text{otherwise,} \end{cases} \tag{5.13}$$

where β is a positive constant called the *loop gain*. Thus, the new-input is formed from the present output where it satisfies the constraints; otherwise from the previous input less a fraction of the present output. The iterative loop for the hybrid input-output algorithm is depicted in Fig. 5.3.

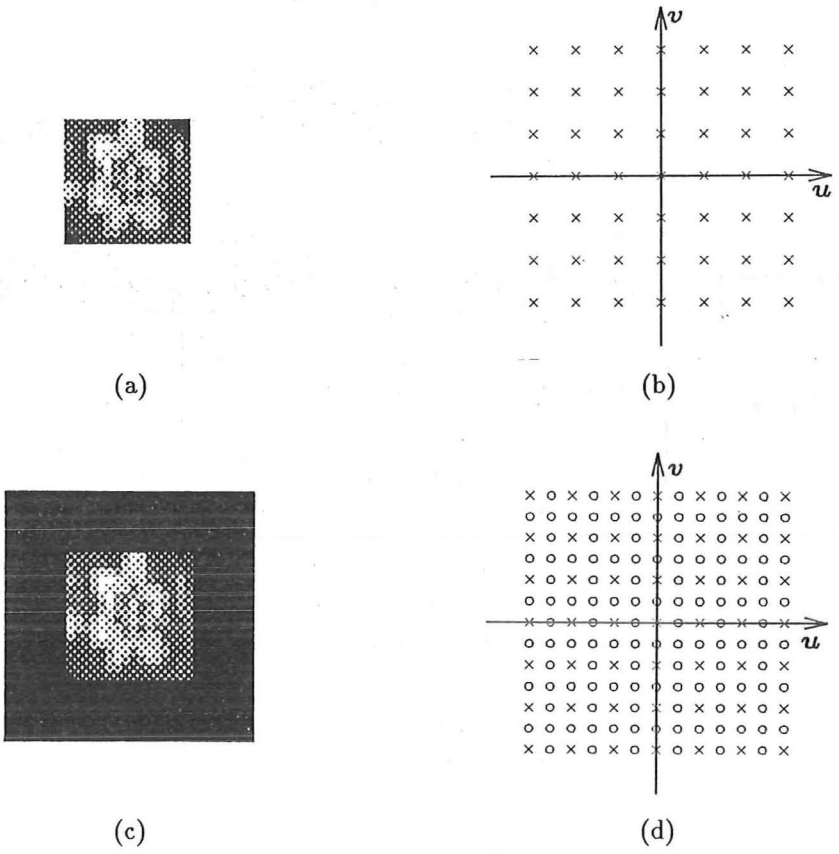
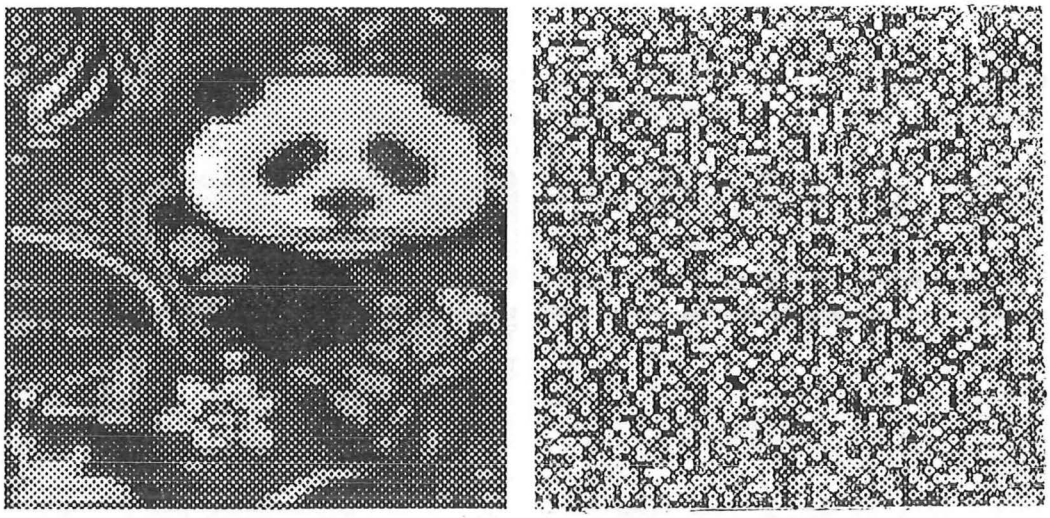


Figure 5.4: Illustration of zero-packing and its effect on the sampling rate in Fourier-space: (a) a 16 × 16 pixel image; (b) an idealized representation of Fourier-space Nyquist sample points; (c) the zero-packed version of (a); (d) Fourier-space sample points of (c). The samples indicated by × correspond to the Nyquist samples of (b), while those marked o indicate the additional sample points due to the zero-packing.

The reasoning behind the hybrid input-output algorithm is somewhat heuristic. It is, however, useful to view the application of the magnitude constraint (i.e. the section of the loop to the right of the dotted line in Fig. 5.3) as a non-linear process. A small change in the input to this process causes a small change in its output, which, to a first order approximation, is proportional to that of the input. Thus, by changing the input appropriately, it is possible to drive the output toward the correct solution. The motivation for replacing (5.9), the image-space constraint of the error-reduction algorithm, with (5.13), is based upon this reasoning.

Examples of Fourier phase retrieval generated by invoking the error-reduction and hybrid input-output algorithms are presented in Fig. 5.6. These reconstructions were produced in the following manner. The true object depicted in Fig. 5.5(a) was *zero-packed* (i.e. the object was inserted into the central portion of a zero-valued image of twice the extent of the original image) to ensure that the oversampling criterion in Fourier-space was satisfied. The concept of zero-packing, and an idealized representation of the Fourier-space oversampling that results, are depicted in Fig. 5.4. Since the true object for the example has an extent of 64 × 64 pixels, the zero-packed image has an extent of 128 × 128 pixels. The magnitude of the spectrum of the resulting image was contaminated to a level of $\epsilon_{\text{dB}} = -30\text{dB}$ and was taken to be the Fourier-space constraint. A buffer filled with pseudo-randomly generated numbers,



(a) (b)

Figure 5.5: The positive 'true' image and the pseudo-random starting image for the example of Fienup's algorithms illustrated by Figs. 5.6 and 5.7.

Figure	Algorithm type	Iterations	Support	Positivity	β
(a)	error-reduction	256	yes	yes	—
(b)	hybrid input-output	32	yes	yes	0.5
(c)	hybrid input-output	64	yes	yes	0.5
(d)	hybrid input-output	256	yes	yes	0.5
(e)	hybrid input-output	256	no	yes	0.5
(f)	hybrid input-output	256	yes	no	0.5

Table 5.3: Details of various strategies employed when generating the reconstructions displayed in Fig. 5.6.

uniformly distributed in the range $0 - 1$, was taken to be the starting image, which is displayed in Fig. 5.5(b)). Details of the image-space constraints applied at each iteration, and the number of iterations applied, for each reconstructions Fig. 5.6(a)–(f), are listed in Table 5.3.

Several points of interest concerning the performance of the error-reduction and hybrid input-output algorithms are clearly revealed in Figs. 5.5–5.7. The first point to note is that the errors resulting from the error-reduction algorithm monotonically decrease (i.e. they never increase). This is apparent from inspection of the graph of E_I versus iteration number displayed in Fig. 5.7(a). Hereafter such a graph is referred to as an *error-curve*. In contrast to error-reduction, the error-curve produced by the hybrid input-output algorithm (Fig. 5.7(b)) fluctuates. No general theoretical analysis of the convergence of the hybrid input-output algorithm has yet been devised and there is no guarantee that its error decreases at each iteration. The fluctuations observed (Fig. 5.7(b)) are typical of those encountered when applying the hybrid input-output algorithm to contaminated data.

Secondly, although error-reduction cannot diverge, it is apparent from the error-curves shown in Fig. 5.7(a) that, after 256 iterations, the algorithm has stagnated. This is confirmed by inspection of Fig. 5.6(a), which bears little resemblance to the

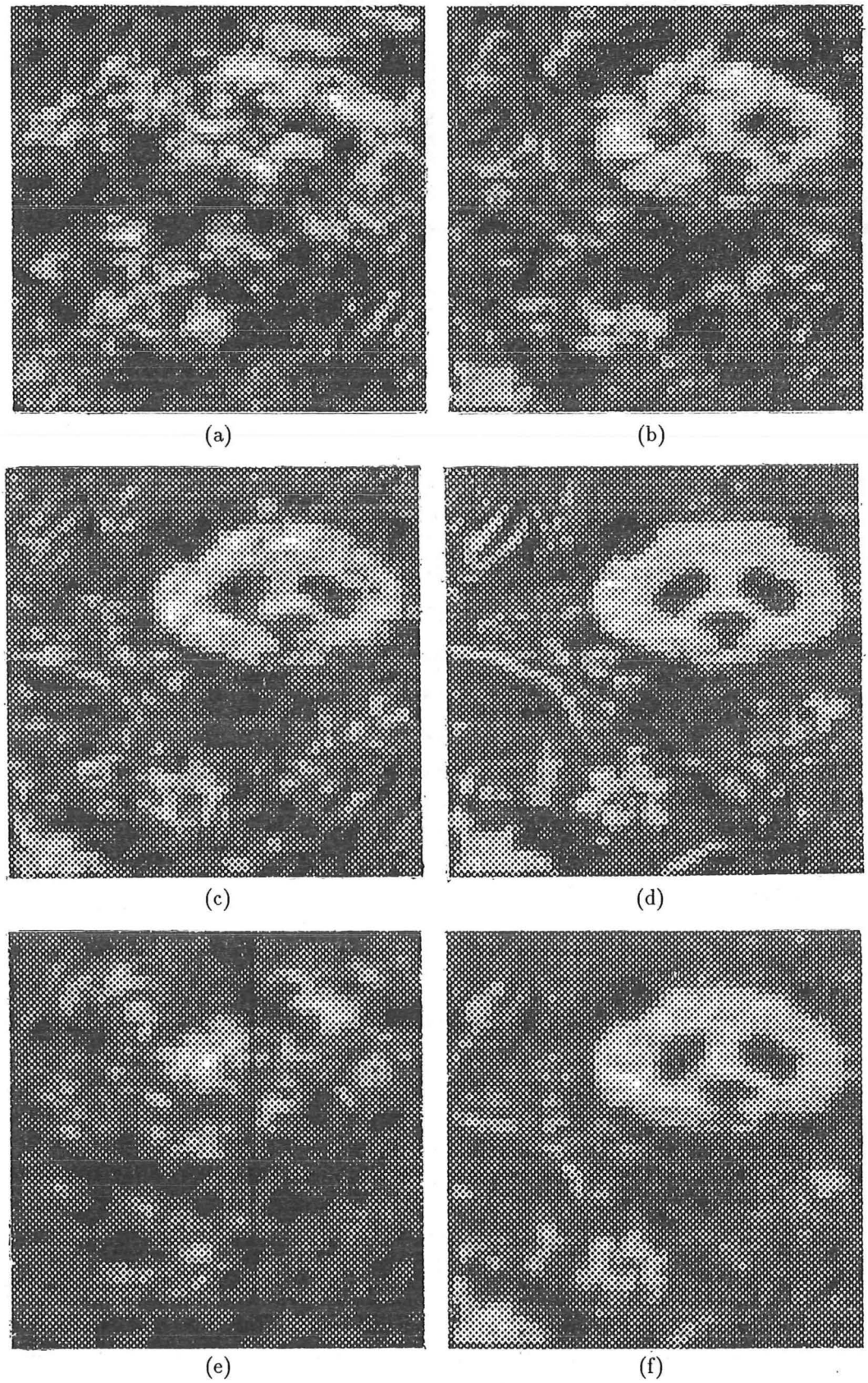


Figure 5.6: Reconstruction of the image shown in Fig. 5.5(a) from the magnitude of its oversampled spectrum by invoking Fienup's iterative algorithms. Contamination to the level of $\epsilon_{\text{dB}} = -30\text{dB}$ was added to the Fourier magnitude and the image depicted in Fig. 5.5(b) was taken as the starting image. Details of each reconstruction are summarized in Table 5.3.

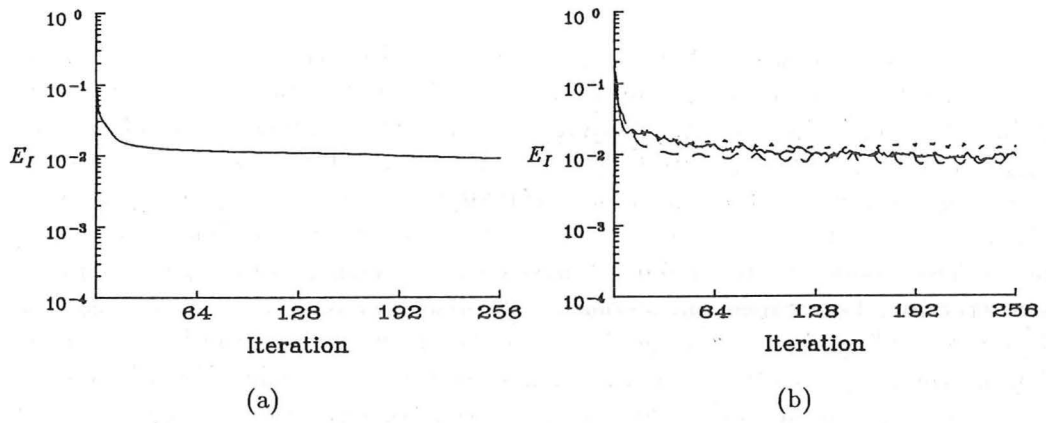


Figure 5.7: Error-curves for the reconstructions presented in Fig. 5.6: (a) error-reduction; (b) hybrid input-output. Solid lines signify application of both support and positivity constraints, dashed lines signify no support constraint, and dotted lines signify no positivity constraint.

true image Fig. 5.5(a). After the same number of iterations the reconstruction from hybrid input-output is significantly more faithful (compare Fig. 5.6(a) and (d)). It is important to note that, although the reconstruction from hybrid input-output is significantly more faithful than for error-reduction, the corresponding image-space error, E_I, is comparable (see Fig. 5.6(b)). This point discussed by Lane (1988, p. 110) who notes that the reconstruction by hybrid input-output for a given value of E_I is significantly superior in visual quality than a reconstruction from error-reduction for the same value of E_I.

Lastly, it appears from Fig. 5.6(e) and (f) that, when the image-space constraints of positivity and support are applied separately, the reconstruction is not as faithful as when these constraints are applied simultaneously. In addition, when no support constraint is applied, there is no necessity for the reconstruction to form centred at the origin of image-space. The reconstruction shown as Fig. 5.6(e) was chosen as the block of 64×64 pixels in the 128×128 reconstruction which contained the greatest energy. The centre of this particular block was located a considerable distance from the origin of image-space.

The problem of stagnation occurs for the hybrid input-output algorithm, although not nearly as regularly, or with as great a severity, as for the error-reduction algorithm. Fienup and Wackerman (1986) and Lane (1988, §5.7) discuss the stagnation problem as it applies to the hybrid input-output algorithm. They propose techniques that effectively overcome the stagnation problem in the majority of practical situations in which it arises.

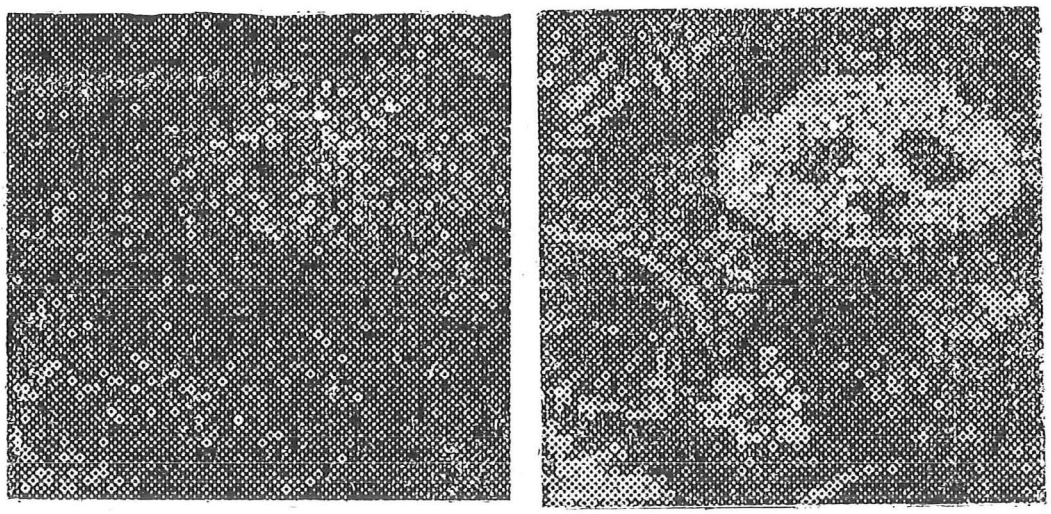
Fienup (1984) discusses a composite technique in which a number of hybrid input-output iterations, are followed by a, usually smaller, number of error-reduction iterations. This procedure is often repeated several times to achieve a reconstruction. Lane (1987) (*cf.* Lane 1988), however, notes that the error-reduction iterations produce a decline in the error measure that is largely illusory, since, when processing returns to the hybrid input-output algorithm, the error returns to its level before the application of error-reduction processing. Furthermore, Lane (1988, §5.3) reports simulations in which the hybrid input-output algorithm, applied alone, converges significantly faster than the composite technique.

M^{c}Callum and Bates (1989) have recently perfected a technique for improving the fidelity of the reconstruction resulting when applying hybrid input-output in the

presence of contamination. Their processing is based on the observation that the convergence of the hybrid input-output algorithm is highly erratic when it processes significantly contaminated data. Because the reconstructed images do not steadily become more faithful as iterations proceed, it is difficult to know when to cease processing. However, M^{c}Callum and Bates (1989) note that the images corresponding to local minima of the error-curve exhibit characteristics of the true image. They find that, if these images are stored and are subsequently intelligently averaged, a superior reconstruction, here called an *averaged reconstruction*, results. The averaging is effected by adding the images together after they have been appropriately aligned. Aligning the images ensures that the versions of the object, which each individual image contain, superimpose in the averaged reconstruction. Processing commences by letting the reconstruction that is deemed to be the 'best' (by some criterion) be the first estimate of the averaged reconstruction. The 'next best' reconstruction is then aligned with the averaged reconstruction before the latter is updated by adding the aligned reconstruction. The alignment is conveniently achieved by correlating the particular reconstruction with the averaged reconstruction. From the maximum of the correlation it is possible to deduce the amount by which the reconstruction must be shifted to align it with the averaged reconstruction. In practice, an additional step is required in the correlation process. Since a phase retrieval algorithm can only hope to recover the image-form of the object, there is no guarantee that the reconstruction of the object at each iteration has the same orientation (it may correspond to the true image rotated by 180°). Thus, in addition to correlating the reconstruction with the averaged reconstruction, it is also necessary to correlate the rotated reconstruction with the averaged reconstruction. If the maximum correlation is encountered with the rotated reconstruction, then it is the rotated reconstruction that is aligned and added to the averaged reconstruction. This sequence of steps continues until all the available reconstructions have been processed. Computer simulations have revealed that the resulting averaged reconstruction is considerably more faithful than the reconstruction corresponding to any minimum of the error-curve (M^{c}Callum and Bates 1989).

M^{c}Callum and Bates (1989) have also investigated methods for improving the computational efficiency of phase retrieval algorithms. They show that a scheme, which operates on increasing numbers of spectral magnitude samples as iterations proceed, leads to significant reductions in the amount of computation required to implement phase retrieval.

The hybrid input-output algorithm has recently been shown to recover the phase of complex data. Fienup (1987) reports reconstructing a complex image having a specialized support, from knowledge of its spectral magnitude. Lane (1987; 1988) finds that, when a sufficiently large number of iterations are applied, the hybrid input-output algorithm can reconstruct a general complex image. An example of a reconstruction of a complex image from its Fourier magnitude, by application of the hybrid input-output algorithm, is depicted in Fig. 5.8. For this simulation the true image was formed by taking Fig. 5.5(a) to be the magnitude, and combining it with a pseudo-random phase, uniformly distributed between $-\pi$ and π. Again the image was zero-packed (Fig. 5.4). The spectral magnitude of the resulting object was extracted to provide the Fourier-space constraint for the hybrid input-output algorithm, while the support constraint comprised the image-space constraint. The starting image was generated by combining a pseudo-random magnitude (Fig. 5.5(b)) with a pseudo-random phase. The reconstructions for 500 and 3000 iterations, with a level of contamination of $\epsilon_{\mathrm{dB}} = -30\mathrm{dB}$ added to the spectral magnitude, are displayed in Fig. 5.8. The principal characteristics of the object are clearly apparent after 3000



(d) (b)

Figure 5.8: Central 64 × 64 pixels of the magnitude of the reconstructed complex image from applying hybrid input-output to the magnitude of the spectrum of the zero-packed true object which was contaminated to a level of $\epsilon_{\mathrm{dB}} = -30\mathrm{dB}$. The true object was generated by taking the magnitude to be Fig. 5.5(a) and combining it with a pseudo-random phase. Reconstruction after: (a) 500 iterations; (b) 3000 iterations.

iterations (see Fig. 5.8(b)). Note also that the reconstructions at 500 iterations is clearly inferior to that for 3000 iterations. The error-curve for this simulation is displayed in Fig. 5.9. A further simulation in which no contamination was added to the spectral magnitude was performed. In this situation the reconstruction after 3000 iterations was indistinguishable from the true image, and the corresponding error was negligible (see Fig. 5.9).

Cederquist *et al.* (1988) have applied Fienup's algorithms to successfully reconstruct a complex image from Fourier intensity data which was generated and recorded in an optical laboratory.

5.3 Direct Phase Retrieval

In recent years considerable theoretical advances have been made in the understanding of the Fourier phase problem. This has led to direct algorithms for the solution of this problem. The arguments underlying direct phase retrieval, and the implications this has for uniqueness, are discussed in this section. The discussion presented here is largely based upon the approach of Lane (1988) (*cf.* Lane *et al.* 1987; Lane and Bates 1987a).

In §5.3.1 it is shown that the zeros of the spectrum of a K-dimensional image form a $(2K-2)$-dimensional surface, henceforth called a zero-sheet, in $2K$-dimensional space. Knowledge of the locations of the zeros of a function is equivalent to knowledge of the function itself. Thus, it is theoretically possible to reconstruct an image from the zeros, apart from a complex scaling constant. Techniques for effecting the reconstruction are introduced in §5.3.2. Furthermore, since the spectrum of a composite image (§2.2) is the product of the spectra of the component images (autocorrelation theorem, Table 2.3), the zero-sheet of the convolution comprises the union of the zero-sheets of the components. It transpires that these individual

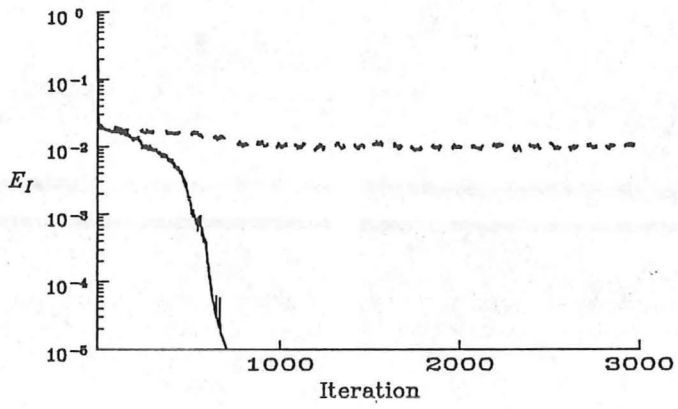


Figure 5.9: Image-space error, E_I, for the reconstruction of a complex image from its spectral magnitude by application of the hybrid input-output algorithm. The solid line corresponds to the contamination free case, while the dashed line shows the error when contamination to a level of $\epsilon_{\text{dB}} = -30\text{dB}$ was added to the spectral magnitude.

zero-sheets are distinct and separable. Techniques for effecting the separation have been realized (Lane and Bates 1987a). The implications that this has for phase retrieval are outlined in §5.3.3. Unfortunately, when contamination is present, the spectrum of a convolution is no longer reducible. Consequently, the zero-sheet of the blurred image no longer comprises distinct and separable zero-sheets, so the sheets of the individual components cannot be distinguished. This deleterious effect of contamination on zero-sheet based direct phase retrieval is mentioned in §5.3.4.

Other authors have reported schemes implementing direct phase retrieval, that are not based on the concept of the zero-sheet. Two of these schemes are mentioned in passing here. Nieto-Vesperinas and Dainty (1986) discuss a technique capable of processing digital (i.e. each pixel is a non-negative integer) uncontaminated images. This scheme is based on factorizing the z-transform of the image. Deighton *et al.* (1985) report a method based on phase closure that recovered a 16×16 image. Lane *et al.* (1987), however, report that this method requires computational effort that increases exponentially with the number of pixels in the image. Therefore, in practice, the technique is only useful for small images.

Napier and Bates (1974) report a two-dimensional phase retrieval algorithm which can be considered a forerunner to the zero-sheet algorithms. Their algorithm is based on reducing the two-dimensional data to one dimension by generating a sequence of one-dimensional projections by invoking the projection theorem (§2.4.1). They note that it is possible to track the complex zeros (§2.8) of the spectrum of the projection of one-dimensional projections thereby effecting phase retrieval.

The complex zeros of the spectrum are the basis of the ensemble blind deconvolution algorithm presented in Chapter 6. The phase retrieval algorithm discussed in §5.3.3 has also been extended to the blind deconvolution problem (Lane and Bates 1987a). This extension is mentioned in §7.1.

5.3.1 Zero-Sheets

The concept of the zero-sheet is described in this subsection with reference to a general two-dimensional compact image $q(x, y)$. The z-transform of this function is

here defined as (see also (2.45))

$$Q(\zeta,\gamma) = \sum_{m,n=0}^{N-1} f_{m,n}\zeta^n\gamma^m, \tag{5.14}$$

where the $f_{m,n}$ are the image-space samples. Compactness of $q(x,y)$ ensures that $Q(\zeta,\gamma)$ is analytic for all finite values of the complex variables (ζ,γ) (see §2.8). Thus, it follows that $Q(\zeta,\gamma)$ is completely characterized by the locations at which its spectrum is zero (see §2.8). Since this property is fundamental to the algorithms discussed in this section, and in Chapter 6, it is convenient to introduce the notation $Z\{Q(\zeta,\gamma)\}$ to denote the set of all points for which the z-spectrum of $g(x,y)$ is zero.

In order to describe $Z\{Q(\zeta,\gamma)\}$, it is convenient to fix one of the variables ζ or γ. By choosing to fix ζ, (5.14) is reduced to

$$Q(\zeta,\gamma) = A(\zeta)\mathrm{e}^{\imath\phi(\zeta)} \prod_{l=1}^{N-1} (\gamma - \gamma_l(\zeta)) \tag{5.15}$$

where $A(\zeta)$ is positive, $\phi(\zeta)$ is real and the $\gamma_l(\zeta)$; $l = 1,\ldots,(N-1)$ are the point zeros corresponding to the particular value to which ζ has been fixed. Now consider the situation in which the variable that has been fixed, ζ in the case of (5.15), is varied infinitesimally. Since $Q(\zeta,\gamma)$ is analytic, and therefore must be continuous and smooth, the positions of the point zeros, $\gamma_l(\zeta)$, necessarily must also shift infinitesimally. Thus, if ζ is varied continuously, the location of the point zeros must necessarily also vary continuously. The variable ζ is complex, meaning that it has two degrees of freedom, so it is possible to vary ζ over a complex plane. As ζ is varied across this plane, it follows that each of the point zeros must describe a two-dimensional surface. This surface is called the *zero-sheet* (Lane *et al.* 1987) of the image $q(x,y)$, or alternatively of the z-spectrum, $Q(\zeta,\gamma)$. Since both ζ and γ are complex, the space that the z-spectrum spans is four-dimensional. However, the zero-sheet is two-dimensional meaning that it is a two-dimensional surface embedded in a four-dimensional space. Lane and Bates (1987a) show that this argument can be extended to a K-dimensional image, in which case the zero-sheet is a $(2K-2)$-dimensional surface embedded in a $2K$-dimensional space (Lane and Bates 1987a). In this chapter, only situations for which $K = 2$ are considered. However, Chapter 6 discusses the zeros-sheets, or zero-maps as they are appropriately called there, for situations in which $K = 1$.

The display of two-dimensional surfaces in a four-dimensional space is not a trivial problem. In order to display a zero-sheet, it is appropriate to display the track of the zeros in the γ-plane as ζ is varied continuously around a closed continuous path (henceforth called a contour) in the complex ζ-plane. A simple method of traversing a contour in the ζ-plane is to specify ζ in polar form (§2.1), i.e.

$$\zeta = |\zeta|\mathrm{e}^{\imath\mathcal{P}[\zeta]}. \tag{5.16}$$

A contour is formed by fixing $|\zeta|$ to a positive constant value, and varying $\mathcal{P}[\zeta]$ continuously from 0 to 2π. A contour in the ζ-plane should, one would expect, lead to a contour in the γ-plane. Thus, by fixing $|\zeta|$, and varying $\mathcal{P}[\zeta]$ through 2π radians, a contour should be traversed on the γ-plane. This often does occur although Lane *et al.* (1987) (*cf.* Lane 1988, §4.5) note that, in practice, the ζ-contour must sometimes be traversed several times in order to close the γ-contour.

Examples of the γ-contours generated from a 32×32 image are plotted in Fig. 5.10. In these examples, as in most examples of zero-contours plotted in this

5.3.2 Image Recovery from Zero-Sheets

In the introductory remarks to this section (§5.3) it was pointed out that it is, in theory, possible to recover an image from its zeros. For a one-dimensional image, this image recovery is a straightforward task since the point zeros are isolated. By multiplying factors formed from the point zeros a polynomial results, i.e.

$$p(z) = \prod_{n=1}^{N-1} (z - z_n), \tag{5.17}$$

where $p(z)$ is the polynomial and the z_n are the point zeros. $p(z)$ is the z-transform of the image, so that the polynomial's coefficients are the samples of the image. This one-dimensional expansion is invoked in Chapter 6, but it is the two-dimensional situation that is of interest in this chapter. Techniques for effecting image recovery for two-or-more dimensions are mentioned here.

When the image is two-dimensional, there are an infinite number of points on its zero-sheet. This considerably complicates the process of reconstructing the image from the zeros since it is necessary to choose the point zeros from which the reconstruction is to be generated. In recent years considerable research has been directed toward recovering a two-dimensional image from its zeros (Curtis and Oppenheim 1987). Two major approaches allowing the image to be recovered from its zero-sheet have been reported (Lane 1988, §4.8).

The first approach (Lane 1988, §4.8.1) involves appropriately selecting the values to which one of the variables (ζ or γ) is fixed. In this manner it is possible to take one-dimensional projections in order to determine the Fourier transform along lines, or slices (see §2.4) in Fourier-space. By making these lines correspond to a regular grid sampled at the Nyquist frequency, the image can be obtained by inverse Fourier transforming (see (2.41)) the image.

The second approach is based on formulating a set of linear equations with the pixels of the image as the unknowns. The technique has recently been reviewed by Curtis and Oppenheim (1987). Their review mainly covers image recovery from the zero crossings of an image, although Izraelevitz and Lim (1987) apply the technique to recover an image from what is essentially a zero-contour. One of the major difficulties with this technique is determining how many zero-sheet samples are required to determine the image-form. Curtis *et al.* (1985) discuss a technique for determining this number based on the number of points two polynomials can have in common. However, Curtis and Oppenheim (1987) note that, provided there are more samples on the zero-sheet than there are pixels in the image, the image-form can be recovered. In addition, the robustness of the technique can be increased by taking more samples of the zero-sheet.

An important point to note, when reconstructing an image from its zeros, is that the resulting image is necessarily only reconstructed to within a complex scaling factor of the original image. It is, in fact, implicit in the thesis that all reconstructions from zeros are subject to this restriction. It should be pointed out, however, that such a restriction is usually of little account, for the reasons discussed in §5.1.1.

5.3.3 Phase Retrieval from Zero-Sheets

Zero-sheet theory has implications for phase retrieval because of the properties embodied in the autocorrelation theorem (see (5.1)) which reveals that solving the phase problem is a sub-problem of the blind deconvolution problem (see the final paragraph

of §5.1.1). An algorithm for implementing phase retrieval, with the concept of the zero-sheet as its basis, is discussed here.

Consider the 2-composite image

$$g(x,y) = f_1(x,y) \odot f_2(x,y) \tag{5.18}$$

and its z-spectrum

$$G(\zeta,\gamma) = F_1(\zeta,\gamma)F_2(\zeta,\gamma). \tag{5.19}$$

Since $G(\zeta,\gamma)$ is the product of $F_1(\zeta,\gamma)$ and $F_2(\zeta,\gamma)$, it follows that $G(\zeta,\gamma)$ is zero whenever either $F_1(\zeta,\gamma)$ or $F_2(\zeta,\gamma)$ are zero. Thus, $Z\{G(\zeta,\gamma)\}$ must be the union of $Z\{F_1(\zeta,\gamma)\}$ and $Z\{F_2(\zeta,\gamma)\}$, i.e.

$$Z\{G(\zeta,\gamma)\} = Z\{F_1(\zeta,\gamma)\} \cup Z\{F_1(\zeta,\gamma)\}. \tag{5.20}$$

Now consider the case of Fourier phase retrieval in which one is given $|Q(u,v)|$ and is required to recover $q(x,y)$. Equation (5.1) reveals that the phase retrieval problem is effectively a deconvolution problem in which $q(x,y)$ is required to be reconstructed from its autocorrelation

$$qq(x,y) = q(x,y) \odot q^*(-x,-y). \tag{5.21}$$

Thus, $q(x,y)$ and $q^*(-x,-y)$ take the part of $f_1(x,y)$ and $f_2(x,y)$, respectively, in (5.18). Consequently the z-spectra of $q(x,y)$ and $q^*(-x,-y)$, i.e.

$$Q(\zeta,\gamma) = \mathcal{Z}[q(x,y)] \quad \text{and} \quad Q^*(\frac{1}{\zeta^*},\frac{1}{\gamma^*}) = \mathcal{Z}[q^*(-x,-y)], \tag{5.22}$$

take the part of $F_1(\zeta,\gamma)$ and $F_2(\zeta,\gamma)$ in (5.20). Thus,

$$Z\{G(\zeta,\gamma)\} = Z\{QQ(\zeta,\gamma)\} = Z\{Q(\zeta,\gamma)\} \cup Z\left\{Q^*(\frac{1}{\zeta^*},\frac{1}{\gamma^*})\right\}. \tag{5.23}$$

Fourier phase retrieval would be achieved if the components of the autocorrelation, $qq(x,y)$ could be recovered. Alternatively, but equivalently, phase retrieval would be realized if the zero-sheet $Z\{QQ(\zeta,\gamma)\}$ could be partitioned into its two constituent zero-sheets, i.e. $Z\{Q(\zeta,\gamma)\}$ and $Z\left\{Q^*(\frac{1}{\zeta^*},\frac{1}{\gamma^*})\right\}$. The techniques discussed in §5.3.2 could then be invoked to recover $q(x,y)$ and $q^*(-x,-y)$. There is no way of knowing which of $q(x,y)$ and $q^*(-x,-y)$ correspond to each of the separated zero-sheets without additional *a priori* information. This fundamental ambiguity in the reconstruction is equivalent to reconstructing the image-form. However, as mentioned in §5.1.1, this is often unimportant since the essential characteristics of an image are contained in its image-form.

Zero-contours of the 5 × 5 pixel autocorrelation of the image specified by Table 5.4(a) are shown in Fig. 5.12. Each of these γ-contours was generated in the same manner as the γ-contours of the image shown in Fig. 5.11. It is apparent that the contours evident in each of Fig. 5.11(a)–(e) are contained in the corresponding plot depicted in Fig. 5.12. However, in addition to these contours of $q(x,y)$, the γ-contours shown in Fig. 5.12 also include the contours of $q^*(-x,-y)$.

Lane *et al.* (1987) describe a practical method for implementing direct phase retrieval by separating the zero-sheets of the components of the convolution. This algorithm is based upon the analytic properties of the z-spectrum, and involves tracking along the zero-sheets in the knowledge that the sheet must be smooth. The algorithm commences by fixing $|\zeta|$ to a constant value, setting $\mathcal{P}[\zeta] = 0$ and

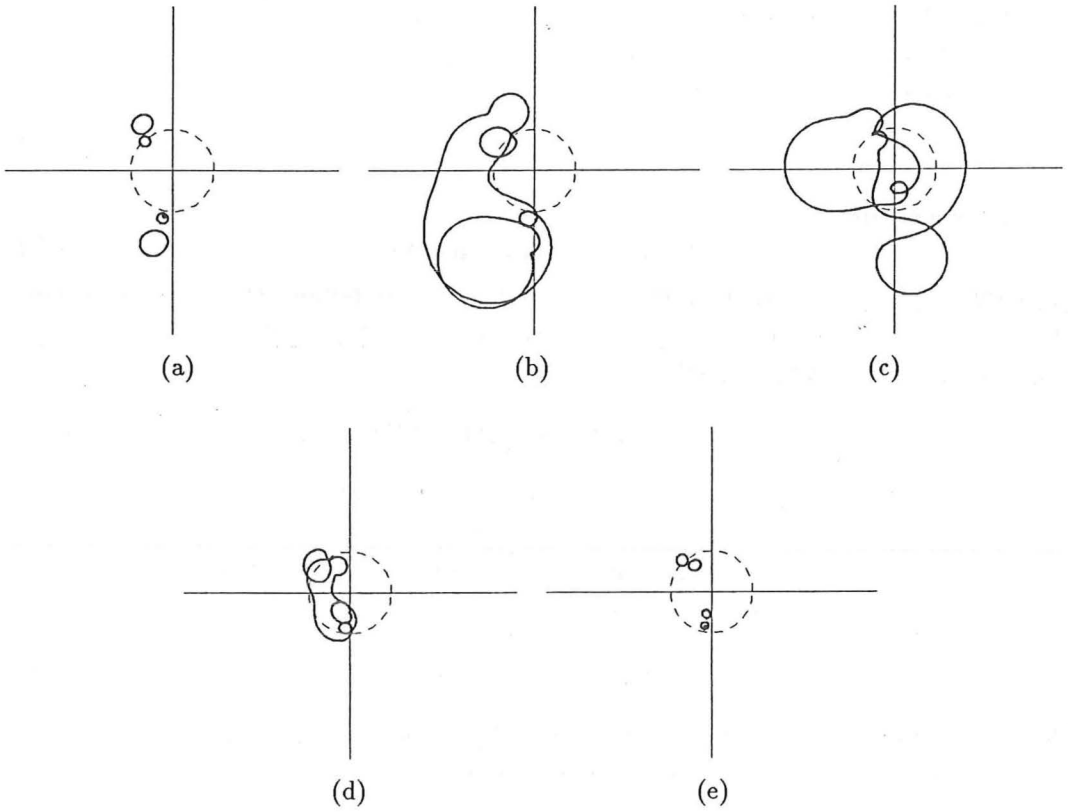


Figure 5.12: Zero-contours constructed from the autocorrelation of the 3 × 3 pixel image described by Table 5.4(a). The values to which ζ was fixed to generate these contours correspond to those invoked in Fig. 5.11.

generating the polynomial described by (5.15). The CPOLY polynomial zero-finding algorithm (Jenkins and Traub 1972) is then invoked to locate the point zeros. Next, $\mathcal{P}[\zeta]$ is incremented by a fixed step and a new polynomial is found. Each of the zeros of this polynomial are then found, not by CPOLY, but with a Newton-Raphson search using the corresponding zero of the first polynomial as the starting point. This method of zero location is more efficient than invoking CPOLY. By repeatedly incrementing ζ, one forces each point zero to track across a portion of the zero-sheet. The increment of ζ is chosen small enough that each point zero's location can be recognized unambiguously from one increment to the next. Lane (1988, §4.7) notes that a more sophisticated approach can be devised by predicting the location of the next zero from the previous two zero positions. This approach has the advantage that it reduces confusion when two zero-sheets are very close, since it helps enforce continuity of the first derivative of the zero-sheet. Since the zero-sheet is analytic, both the zero-sheet, and its first derivative must be continuous. Consequently, this approach allows two intersecting zero-sheets to be unambiguously separated. This ability to track through an intersection is indicated by the arrows in Fig. 5.13(a).

Lane *et al.* (1987) present an example of recovering the phases of ten different complex images, each comprising 16 × 16 pixels, from their 31 × 31 pixel autocorrelations. They also note that techniques based upon separating zero-sheets have the advantage of being capable of processing images that are in general complex. However, if it is known *a priori* that an image is real, the computational burden can be reduced by invoking symmetry properties of the zero-sheet.

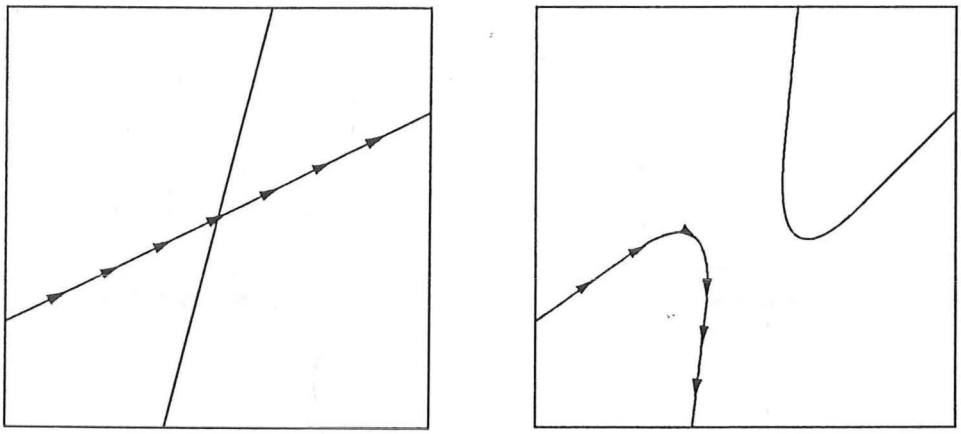


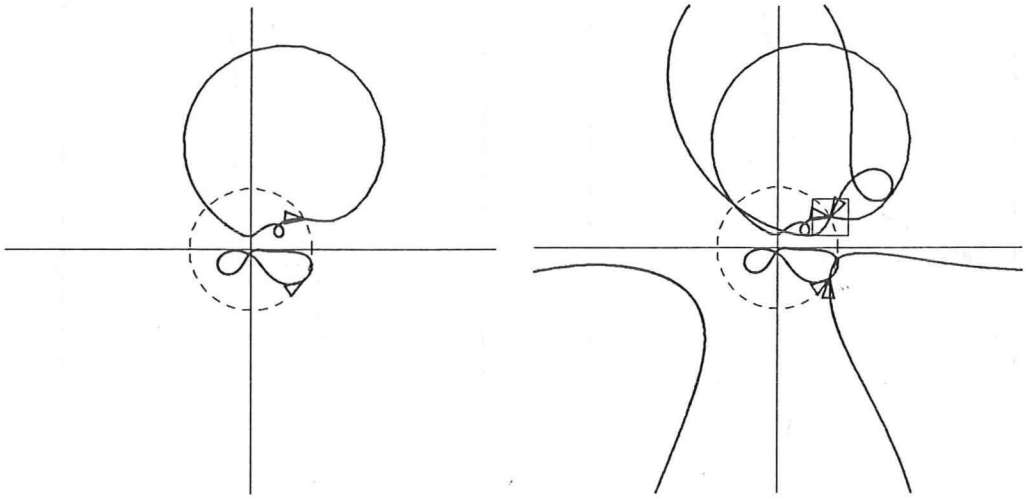
Figure 5.13: An idealized representation of the bridges induced between the zero-sheets when contamination is added to a convolution: (a) an intersection of zero-contours with no contamination; (b) bridges between the contours have formed when contamination is added. The arrows superimposed upon the contours represent the path followed by the algorithm that tracks the contour based on the assumption of analyticity of the zero-sheet. When contamination is present (see (b)) this path follows the bridge meaning that the component zero-sheets can not be separated.

5.3.4 Effect of Contamination on Zero-Sheets

When a two-dimensional convolution is contaminated, it can no longer be exactly deconvolved into smaller images. Because contamination is always present in practice, an exact convolution almost never arises. In mathematical parlance, it is said that the set of images that are convolutions form a set of measure zero (Hayes and M^{c}Clellan 1982). Thus, when an image that is reducible is perturbed slightly by the addition of contamination, the resulting image is, essentially always, irreducible (Sanz and Huang 1985).

Consider a two-composite image $g(x, y)$ and its zero-sheet as defined by (5.18) and (5.20) respectively. Consider in particular the behaviour of the zero-sheet when contamination is added to the image. Since a contaminated image is no longer reducible, the zero-sheet no longer comprises the two distinct zero-sheets of the components. Instead, these zero-sheets link to form a single zero-sheet. Analyticity of the z-spectrum ensures that a small perturbation of the image causes a small perturbation of the z-spectrum. Thus, if the level of contamination is small, the zero-sheet is only slightly distorted. In order for an infinitesimal amount of contamination to induce linkage of the component zero-sheets, it follows that they must intersect. Walker (1950, p. 40) notes that, in general, two M-dimensional surfaces, existing in an N-dimensional space, intersect in a space of dimension $2M - N$. Thus, two-dimensional zero-sheets in four-dimensional space would be expected to intersect in a zero-dimensional space, i.e. at discrete points. This is consistent with the linkage of zero-sheets induced by contamination. Bates and Lane (1988) (*cf.* Lane 1988, §4.9) call the links formed between the zero-sheets *bridges*. An idealized representation of bridging is depicted in Fig. 5.13.

The existence of such a linkage between zero-sheets, caused by the presence of contamination, means that the direct phase retrieval algorithm discussed in §5.3.3



(a) (b)

Figure 5.14: Zero-contours generated with $|\zeta| = 1.0$ and varying $\mathcal{P}[\zeta]$ through 2π radians for: (a) $q(x,y)$, the 3×3 image whose pixels are listed in Table 5.4(b) and; (b) $qq(x,y)$, the 5×5 autocorrelation of this image. $Z\{Q(\zeta,\gamma)\}$ and $Z\left\{Q(\frac{1}{\zeta^*},\frac{1}{\gamma^*})\right\}$ intersect at $\gamma = \frac{\sqrt{3}}{2} + \imath\frac{1}{2}$, when $\zeta = 1.0e^{\imath 0.0}$. The square, centred about this point on (b), delineates the region of the γ-plane that is enlarged in each of the plots shown in Fig. 5.15. Arrows, superimposed on contours at the point for which $\mathcal{P}[\zeta] = 0.0$, identify the direction of the contour for increasing $\mathcal{P}[\zeta]$.

is no longer able to effect a separation of the zero-sheets. This is because the algorithm tracks the zero-sheet on the assumption that the zero-sheet is smooth. In the absence of contamination the algorithm can track across intersections of zero-sheets because any abrupt change in the sheet would violate the requirement that the zero-sheet be analytic. Tracking over an intersection is indicated by the arrows shown in Fig. 5.13(a). However, in the presence of contamination the zero-tracking procedure follows the bridge, as shown in Fig. 5.13(b). Thus, the phase retrieval algorithm described in §5.3.3 can not, in general, recover the spectral phase in the presence of contamination.

A pictorial study of the effect of contamination on the zero-sheets of an autocorrelation is depicted in Figs. 5.14 and 5.15. The pixels of the 3×3 image in this example are listed in Table 5.4(b) and the γ-contours of the image for $|\zeta| = 1.0$ are displayed in Fig. 5.14(a). The corresponding γ-contours of the image's autocorrelation are shown in Fig. 5.14(b). Note that arrows have been added to the contours to identify the point at which $\zeta = 1.0e^{\imath 0.0}$. The direction of the arrow indicates the path of the contour for increasing $\mathcal{P}[\zeta]$. The image was constructed so as to ensure that the zero-sheets of the components of the autocorrelation, i.e. $Z\{Q(\zeta,\gamma)\}$ and $Z\left\{Q(\frac{1}{\zeta^*},\frac{1}{\gamma^*})\right\}$, intersect at the points $\gamma = \frac{\sqrt{3}}{2} + \imath\frac{1}{2}$ and $\gamma = \frac{\sqrt{3}}{2} - \imath\frac{1}{2}$, when $\zeta = 1.0e^{\imath 0.0}$. Inspection of Fig. 5.14(b) confirms that the sheets do intersect at these points since the arrows on the two contours meet there. A detailed pictorial study of contamination induced distortion of the zero-sheets in the vicinity of the intersection of the zero-sheets comprising the autocorrelation is presented in Fig. 5.15. Each of these plots is an enlargement of the square shown in Fig. 5.14(b), centred about the point $\gamma = \frac{\sqrt{3}}{2} + \imath\frac{1}{2}$, and of size (0.6,0.6). The plots comprising the three columns,

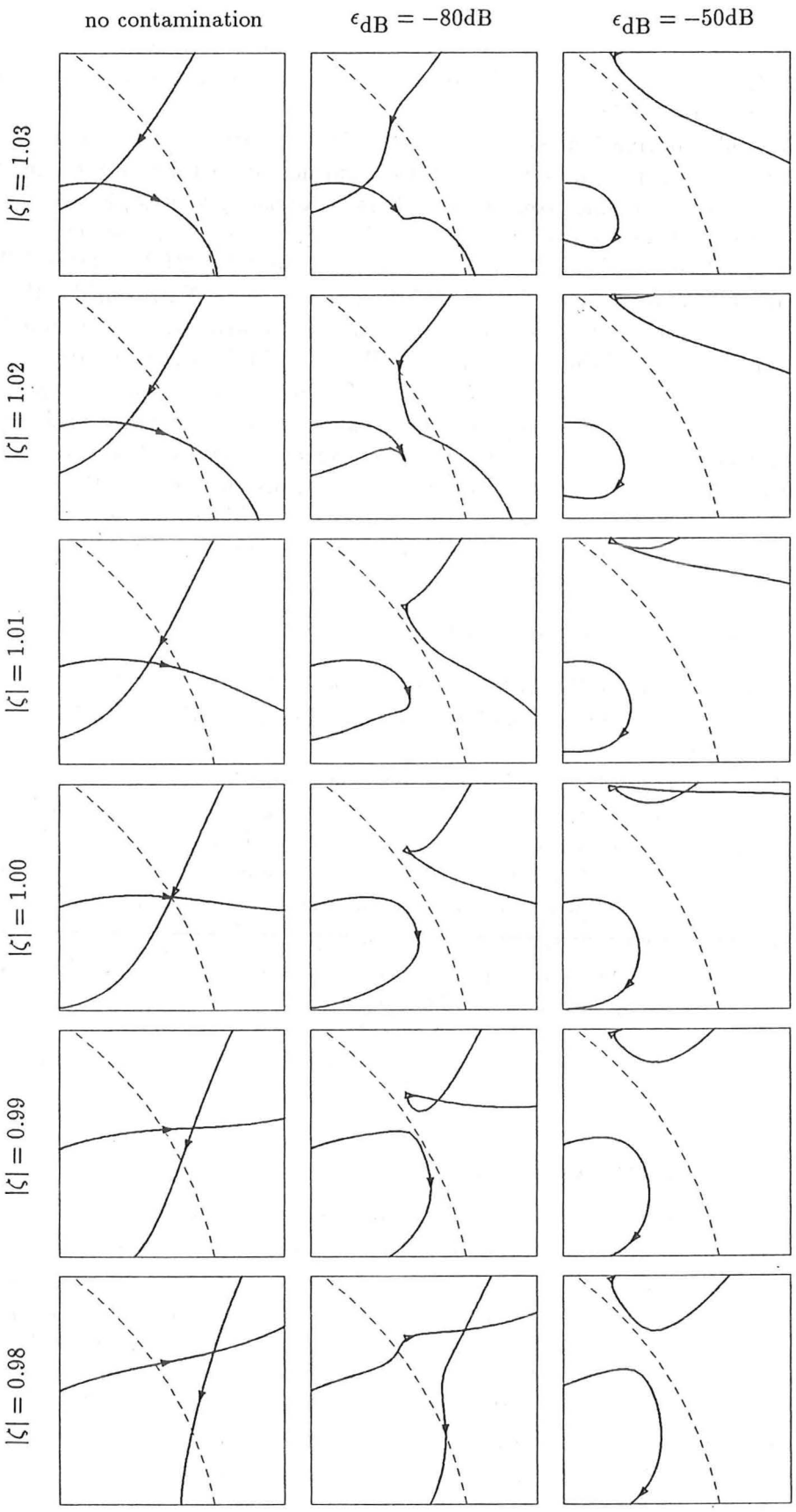


Figure 5.15: Contamination induced distortion of a zero-sheet of the autocorrelation of the image listed in Table 5.4(b) near an intersection of the component zero-sheets. Plots in each column correspond to the same contamination level, and in each row to the same value of $|\zeta|$ (the actual values are indicated on plots). Refer also to the caption of Fig. 5.14.

from left to right, correspond to levels of contamination of: none; $\epsilon_{dB} = -80$dB; $\epsilon_{dB} = -50$dB. The six rows from top to bottom correspond to six equispaced values of $|\zeta|$ in the range 1.03–0.98.

Considerable information about the nature of the distortion at an intersection of a pair of of zero-sheets is apparent in Fig. 5.15. Consider, for instance, the progression of contours shown in the first row of plots. It is apparent that, as $|\zeta|$ is varied from 1.0, the arrows no longer meet, meaning that the zero-sheets no longer intersect in four-dimensional space. Since the contours only intersect for $\mathcal{P}[\zeta] = 0.0$ it follows that the component zero-sheets do indeed intersect at points. Now consider the row corresponding the $|\zeta| = 1.0$. The first plot in this row, for no additive contamination, shows the intersection of the zero-contours. However, bridges have clearly formed in the presence of even a very small amount of contamination (column two). An increased contamination level (column three) causes even more severe bridging, in that the contours have drifted even further from the location of the true intersection. Several other characteristics in the presence of noise are of interest. For instance, it seems that 'loops' in the contours often form on the bridges near to the point at which $\mathcal{P}[\zeta] = 0.0$. However, for a given pair of bridges, only one has a loop.

5.4 Magnitude from Phase

Several authors including Hayes (1982) and Lane and Bates (1987b), discuss the practical importance of algorithms for recovering the spectral magnitude from the phase (see §5.1.2).

Lane (1988) divides the magnitude problem into two classes. The first class, the *pure magnitude problem*, occurs when complete information about the phase of the image is known. The second, the *modified magnitude problem*, occurs when the phase is only known modulo π. This latter problem presents significantly greater difficulties than the pure magnitude problem because considerably less information about an image is provided. Consider the reconstruction Fig. 5.16, generated by combining a unity magnitude and the modulo π phase of the spectrum of Fig. 5.1(a). It is obvious that this reconstruction bears no resemblance to the original image. Comparing this figure and the corresponding figure generated using the true phase (Fig. 5.1(b)) suggests that the modulo π phase preserves significantly less information about the true image.

5.4.1 Pure Magnitude Problem

Hayes *et al.* (1980) (*cf.* Oppenheim and Lim 1981, §4; Hayes 1982) describe an algorithm for implementing a solution to the magnitude problem based on the general iterative processing loop shown in Fig. 5.2. When this loop is used to implement magnitude retrieval the Fourier-space constraint is derived from the known phase. The image-space constraints (§5.2) comprise a support constraint, and reality and positivity if appropriate.

As with the phase problem, the Fourier-space data must be sampled at a rate greater than, or equal to, twice the Nyquist rate. Thus, for the examples presented in this section, the image was zero-packed (Fig. 5.4) before its spectral phase was extracted to form the Fourier-space constraint.

Significant differences exist between the phase and magnitude problems. Firstly the location and orientation of an object are contained in the spectral phase. Thus, the notion of image-form, introduced when discussing the phase problem, is inappropriate for the magnitude problem. Furthermore, it follows that unless the support

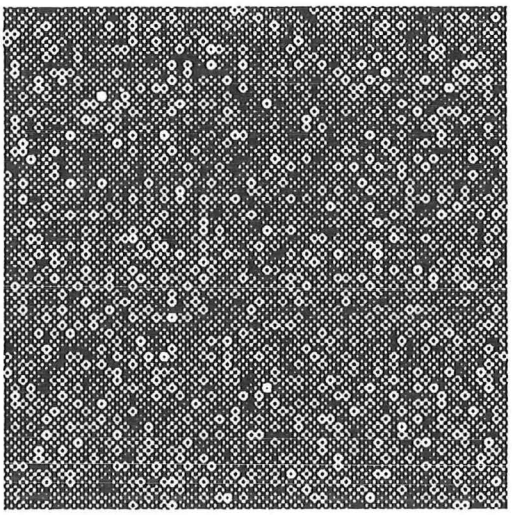


Figure 5.16: Reconstruction from the modulo π phase of the image shown in Fig. 5.1(a), combined with unity magnitude. It is apparent, when comparing this reconstruction with that displayed in Fig. 5.1(b), that considerably less information about the true object is preserved in the modulo π phase than in the true phase.

is correctly positioned, convergence to the true solution cannot occur. Secondly, the magnitude of an image is not preserved in its spectral phase, so the magnitude retrieval algorithms can only reconstruct an image to within a real scaling constant. Thus, the concept of a *normalized image* is invoked in this thesis to describe the reconstruction generated by a magnitude retrieval algorithm. Thirdly, the support of the image is not deducible from its spectral phase, as it is from its spectral magnitude. Derivation of the support from the magnitude is possible, uniquely for a positive image, by invoking the autocorrelation theorem (Table 2.3) and the extent of autocorrelation theorem (2.28).

The inability to deduce the true support from the given phase is compounded by the multiple solutions that arise from magnitude retrieval algorithms when the available estimate of the support is larger than the true support. Such a support is here called *overestimated.* Since the Fourier-space constraint requires only that the spectral phase of the reconstruction be zero, it follows that the convolution of the true image with any function having zero spectral phase satisfies this constraint. The extent of convolution theorem (§2.2) states that the extent of a convolution in each coordinate direction is the sum of the extents of the components of the convolution in that coordinate direction. Thus, the extent of a convolution is necessarily larger than the support of each of the convolution's components. Applying this insight to the magnitude problem, it follows that if the exact support of the true image is known, then the only solution to which the algorithm can converge is the true solution. However, consider the situation in which the support is overestimated. Now, the reconstructed function can be the convolution of the true object with a computationally induced psf whose support is sufficiently small that the extent of convolution theorem is satisfied. Thus, when the support is overestimated, the reconstruction is of the form

$$\hat{q}(x,y) = q(x,y) \odot s_i(x,y), \tag{5.24}$$

where $\hat{q}(x,y)$ denotes the alternative solution to the magnitude problem and $s_i(x,y)$ denotes the computationally induced psf existing at the i^{th} iteration. In addition,

to satisfy the Fourier-space constraint that $\mathcal{P}\left[\hat{Q}(u,v)\right] = \mathcal{P}[Q(u,v)]$, $\mathcal{P}[S_i(x,y)]$ must be zero. In order to have zero phase $s_i(x,y)$ must be symmetric, and also have a dominatingly large positive delta function at the origin. An example of such a computationally induced psf is shown in Fig. 5.17(d). Since this delta function dominates $s_i(x,y)$, the blurring in the reconstruction is not severe.

The blurring induced by the computationally induced psf in the pure magnitude problem is apparent in Fig. 5.17(c) and Fig. 5.18(d). Lane (1988, §6.1) notes that, in practice, the reconstruction always fills the extent of the estimated support. Thus, the reconstruction always corresponds to $\hat{q}(x,y)$ rather than $q(x,y)$. Consequently, the iterative loop only converges to the correct solution in situations where the support is known exactly.

Bates and Lane (1988) note that if the extent of the object is unknown, but it is known that the true support is the minimum compatible with the given phase (which can almost always be expected to be the case), then a useful procedure for estimating the support is as follows. The size of the image-box is chosen as large as is compatible with the given samples of $\mathcal{P}[F(u,v)]$. The magnitude retrieval algorithm is then applied with successively smaller image-boxes until the procedure no longer converges, indicating that the image-box is too small. It follows that the true image-box is the smallest that yields convergence. Lane (1988) reports computer simulations that suggest that, even in the presence of significant amounts of contamination, this procedure yields useful estimates of the true supports.

It appears that the iterative loop for the magnitude problem always converges (Youla and Webb 1982). However, the convergence is often slow when dealing with large or complex images. Hayes *et al.* (1980) (*cf.* Oppenheim *et al.* 1982; Lane 1988) and Levi and Stark (1984) describe methods for improving the convergence rate of the iterative algorithm. Fortunately, the stagnation problem, apparent in similar iterative approaches to solving the phase problem (see §5.2), does not appear to arise in the magnitude problem.

It is appropriate to add, as a final note in this discussion of the pure magnitude problem, that not all images can be uniquely recovered from knowledge of their spectral phase. Consider an image that has zero spectral phase. If such a phase is applied to a magnitude retrieval algorithm, any function that is symmetric and has a sufficiently dominant delta function at the origin satisfies the Fourier-space constraint. Since there are infinitely many such functions, no unique solution can be obtained. Furthermore, consider the outcome of invoking the support determining procedure, due to Bates and Lane (1988), to such an image. This procedure, which is mentioned earlier in this subsection, involves repeatedly applying the magnitude retrieval algorithm, each time reducing the extent of the support until the magnitude retrieval algorithm no longer converges. Because in this situation the spectral phase is zero, it follows that minimization of the support would result in the recovery of a delta function. The corresponding estimate of the support would, thus, be one pixel by one pixel, regardless of the actual extent of the image.

5.4.2 Modified Magnitude Problem

An iterative algorithm for solving the modified magnitude problem (i.e. when the phase is only given modulo π) has been developed by Lane and Bates (1987b) (*cf.* Lane 1988, §6.2). This algorithm is similar to the iterative algorithm discussed in §5.4.1 for implementing the pure magnitude problem, although the Fourier-space constraints are somewhat different.

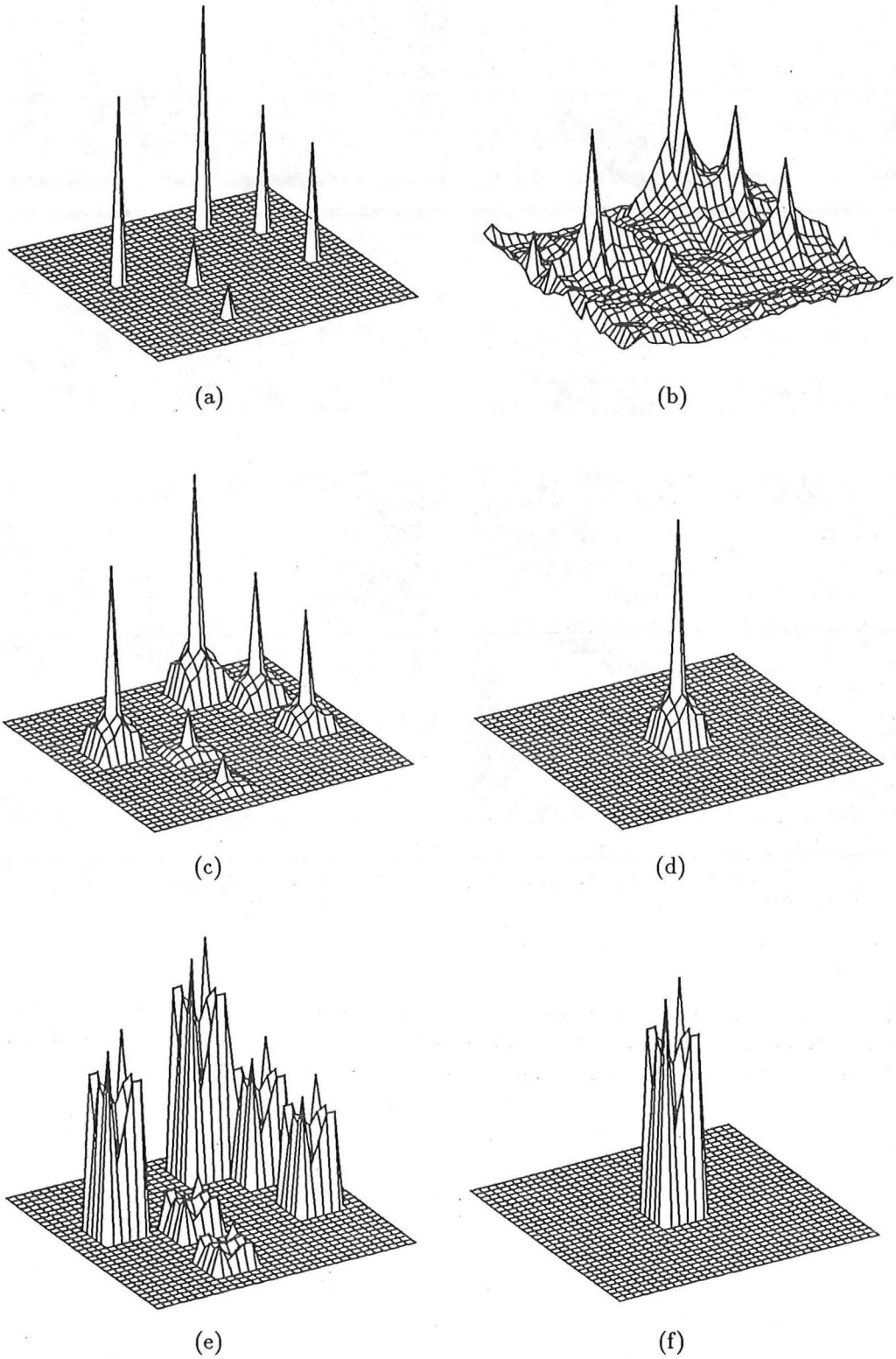


Figure 5.17: Retrieval of image from its spectral phase: (a) true object $q(x, y)$ of extent 22×22 pixels; (b) reconstruction with too small a support (20×20 pixels); (c) reconstruction from true Fourier-space phase with overestimated support (29×29 pixels); (d) computationally induced psf; (e) reconstruction from modulo π phase with overestimated support (29×29 pixels); (f) computationally induced psf.

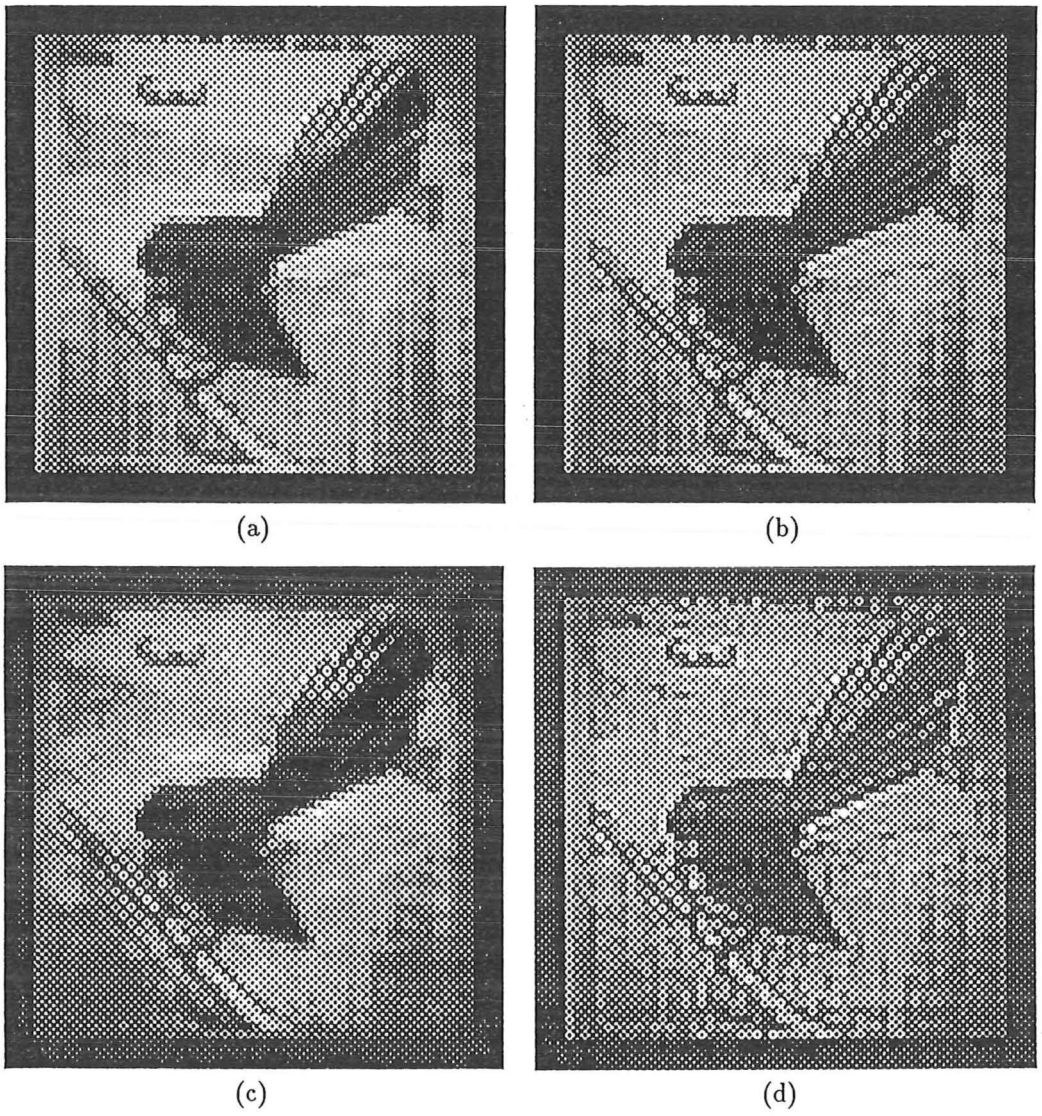


Figure 5.18: Retrieval of image from spectral phase: (a) true image (of support 56 × 56 pixels); Reconstructions after 100 iterations with: (b) exact support; (c) underestimated support (54 × 54 pixels); (d) overestimated support (64 × 64 pixels).

Since the phase is only known modulo π, it is necessary to determine whether

$$\mathcal{P}[F(u,v)] = \mathcal{P}[F(u,v)] \text{modulo } \pi \tag{5.25}$$

or

$$\mathcal{P}[F(u,v)] = (\mathcal{P}[F(u,v)] \text{modulo } \pi) + \pi. \tag{5.26}$$

Thus, the application of the Fourier-space constraint must be modified to select whether (5.25) or (5.26) is chosen as the true phase. Lane (1988, §6.2) notes that the simplest approach is to assume that the required phase is given by whichever of (5.25) or (5.26) is 'closest' to $\mathcal{P}[K_i(u,v)]$. This selection is usefully implemented as (Lane and Bates 1987b)

$$\text{If } |\mathcal{P}[F(u,v)] - \mathcal{P}[K_i(u,v)]| \begin{cases} \leq \frac{\pi}{2} & \text{then} \quad \mathcal{P}[K_i'(u,v)] = \mathcal{P}[F(u,v)] \\ > \frac{\pi}{2} & \text{then} \quad \mathcal{P}[K_i'(u,v)] = \mathcal{P}[F(u,v)] + \pi \end{cases} \tag{5.27}$$

The most significant difference between the pure and modified magnitude problems lies in the differences of the form of the symmetric computationally induced psf when the support is overestimated. In the pure magnitude problem the symmetric psf must be of zero phase, meaning that it must be symmetric and have a dominatingly large delta function at the origin. The presence of such a delta function ensures that the blurring caused by this psf is usually not too severe. However, in the modified magnitude problem the psf need only be symmetric, which follows since the spectral phase of a symmetric function can have only two values, 0 or π, both of which correspond to zero modulo π. Because the psf no longer needs a dominatingly large delta function, the blurring is almost always more severe in reconstructions generated by modified magnitude retrieval algorithms than in those generated by pure magnitude retrieval algorithms. The differences in typical forms of the psfs for the two cases are illustrated by Fig. 5.17(d) and (f). The latter psf, which is induced by the modified magnitude retrieval algorithm, clearly does not have a large central delta function like that evident in Fig. 5.17(d). The blurring apparent in Fig. 5.17(e), caused by the psf shown in Fig. 5.17(f), is also significantly more severe than that apparent in Fig. 5.17(c).

A further comparison of the blurring induced by the pure and modified magnitude problems is provided by Figs. 5.18 and 5.19. When the support is estimated exactly, the reconstructions from both algorithms are comparably faithful (compare Fig. 5.18(b) and Fig. 5.19(b)). However, when the support is overestimated the reconstruction from the pure magnitude problem (Fig. 5.18(d)) is significantly more faithful than that from the modified magnitude problem (Fig. 5.19(d)).

Images that are symmetric can not be uniquely recovered from their modulo π phase. This follows as an extension of the argument presented in the final paragraph of §5.4.2, since the spectral phase of such an image is zero modulo π.

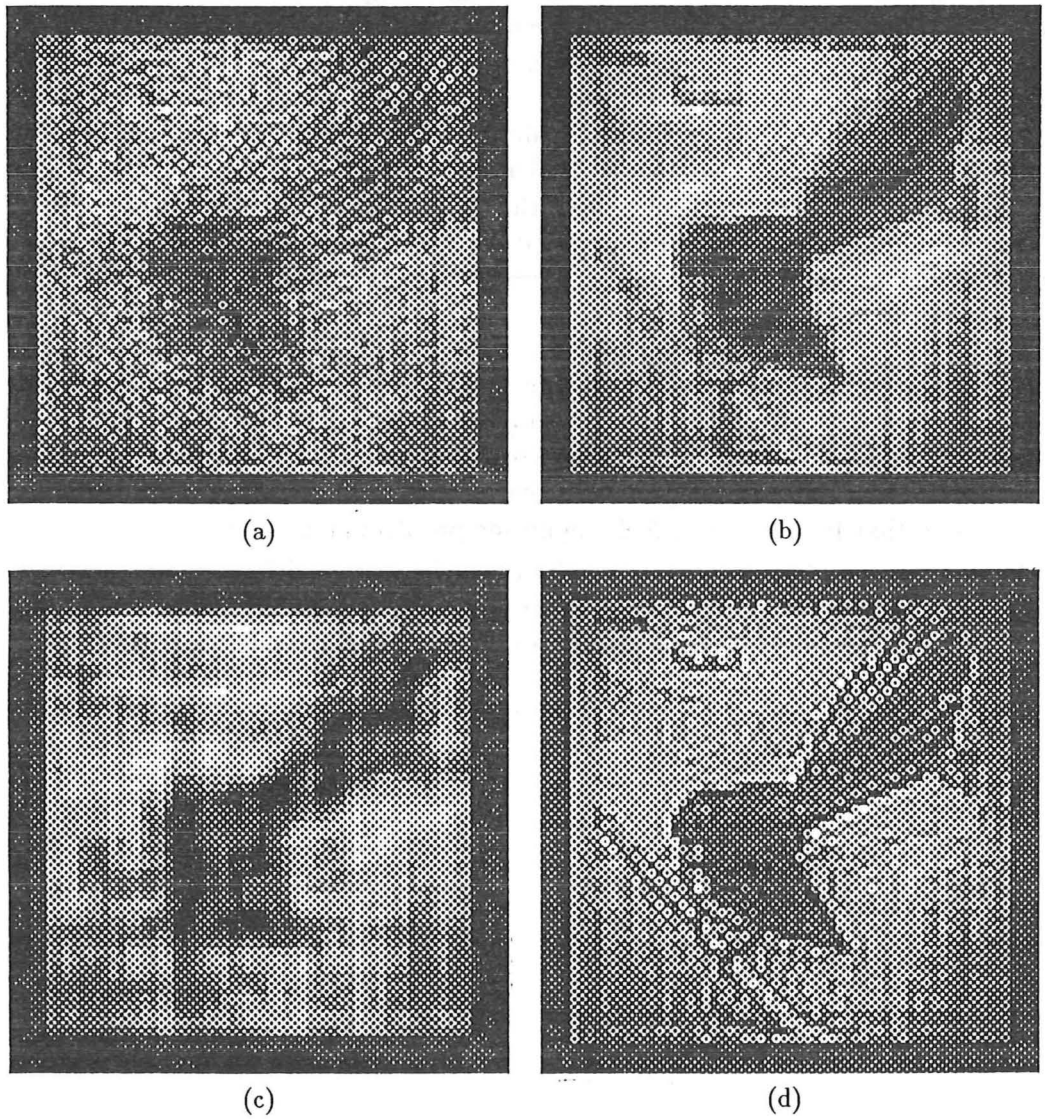


Figure 5.19: Reconstruction of the object shown in Fig. 5.18(a) from the modulo π phase of its spectrum: (a) after 10 iterations with exact support (54 × 54 pixels); (b) after 100 iterations with exact support; (c) after 100 iterations with too small a support (54 × 54 pixels); (d) after 100 iterations with too large a support (64 × 64 pixels).

Chapter 6

Zero-and-add

A new deconvolution technique known as ZAA (Davey *et al.* 1986) is presented in this chapter. This technique was initially developed to complement SAA in the astronomical setting (Bates *et al.* 1985; Sinton *et al.* 1986) by providing a consistent method of countering the ghosting that appears in the basic SAA image of certain classes of objects (see §4.8). ZAA is, however, a speckle imaging technique in its own right (Davey *et al.* 1986; Sinton 1986).

ZAA is an ensemble blind deconvolution technique (see §1.5) in that it recovers a general object from an ensemble of differently blurred versions of that object. The ZAA principle employs theory relating to the set of complex zeros (i.e. the zero-sheet), of the spectrum of an image. Thus, §§2.8 and 5.3 serve as general introductions to ZAA. Additional preliminaries, specific to the ZAA technique, are presented in §6.1.

The application of ZAA to one-dimensional speckle images is described in §6.2. For clarity §6.2.2 describes the application of the technique to uncontaminated speckle images. Methods for countering the deleterious effects of contamination are introduced in §6.2.3. Further details of the ZAA technique are given by Bates *et al.* (1985), Sinton *et al.* (1986), Davey *et al.* (1986) and Sinton (1986).

A new composite scheme for processing one-dimensional infrared speckle images (see §4.9) has recently been perfected (Davey *et al.* 1988a). This technique, which is discussed in §6.3, involves generating several SAA images, each with a different SAA psf. ZAA is then invoked on these SAA images to extract the true object. This processing not only deconvolves the SAA psfs, but is also capable of compensating for pseudo-random contamination introduced during capture of the speckle images.

In order to be generally applicable, any astronomical speckle imaging technique must be capable of processing two-dimensional speckle images. Extensions to ZAA which enable it to operate on two-dimensional data are presented in §6.4. The first extension utilizes the projection theorem §2.4.1, while the second (Bates and Lane 1987b) has its origins in zero-sheet theory (§5.3.1) (Lane *et al.* 1987; Lane and Bates 1987a).

The usage of complex zeros in ZAA is related to the use of zeros in self-consistent deconvolution (Bates *et al.* 1976; M^{c}Kinnon *et al.* 1976), astronomical imaging (Bates and Napier 1972; Walker 1981; Bates and Fright 1982), the Fourier phase problem (Bates 1978; Fright 1984; Lane *et al.* 1987; Lawton and Morrison 1987; Lane 1988) and general blind deconvolution (Lane and Bates 1987a). Napier (1971, §5.2) presents a useful summary of the applications of complex zero theory.

6.1 Preliminaries

6.1.1 The Zero-and-Add Principle

ZAA is based on the principle of the *zero-map* (Bates *et al.* 1985; Davey *et al.* 1986) which is the specialization of the zero-sheet introduced in §5.3.1 for $K = 1$. In §5.3.1 it is noted that the zero-sheet of a K-dimensional object is a $(2K - 2)$-dimensional surface existing in $2K$-dimensional space. So, when $K = 1$, the zero-sheet is zero-dimensional, implying that the zero-map comprises unconnected (or discrete) points on a complex plane.

Consider now the task of deconvolving the function $g(\mathbf{x})$, which can be written as the convolution of two functions $f(\mathbf{x})$ and $h(\mathbf{x})$ i.e.

$$g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x}) \longleftrightarrow G(\mathbf{w}) = F(\mathbf{w})H(\mathbf{w}), \tag{6.1}$$

where $\mathbf{w}$ is a complex position vector (*cf.* (2.47)). Since $G(\mathbf{w})$ is the product of $F(\mathbf{w})$ and $H(\mathbf{w})$, $G(\mathbf{w})$ is zero whenever either (or both) of $F(\mathbf{w})$ or $H(\mathbf{w})$ is zero. Invoking the terminology $Z\{G(\mathbf{w})\}$ (see §5.3.1) to denote the set of zeros, or zero-sheet, of the spectrum of $g(\mathbf{x})$, it follows that $Z\{G(\mathbf{w})\} = Z\{F(\mathbf{w})\} \cup Z\{H(\mathbf{w})\}$ (*cf.* (5.20)), where $\cup$ is the set union operator. It is useful here to recall that an image is completely specified (apart from an arbitrary complex scaling constant) by the positions at which its spectrum is zero (refer to §§2.8 and 5.3.3). Thus, if one could partition the set of zeros comprising $Z\{G(\mathbf{w})\}$ into the two constituent sets $Z\{F(\mathbf{w})\}$ and $Z\{H(\mathbf{w})\}$, $g(\mathbf{x})$ would effectively be deconvolved. In two-or-more dimensions, analyticity of the image ensures that the zeros form a continuous surface. Furthermore, when an image is a convolution, the zero-sheets of its components comprise the convolution's zero-sheet. In the absence of contamination these zero-sheets are necessarily distinct. This allows one to track the zeros over the surface, thereby partitioning the zero-sheet of the convolution into two constituent zero-sheets (this is described in §5.3.1). Unlike for the more-than-one-dimensional situation, when $K = 1$ the zeros no longer form a continuous surface, since the zero-map is zero-dimensional. Consequently, it is no longer possible to partition $Z\{G(w)\}$ without extensive *a priori* information.

Consider now the situation in which one is presented with an ensemble $\{g_m(x) = f(x) \odot h_m(x)\}$, for M values of the integer index m, of differently blurred versions of the one-dimensional object, $f(x)$. The task of recovering $f(x)$ from the ensemble of blurred images is an ensemble blind deconvolution problem (*cf.* §1.5). Given that one can compute the zero-map of each $g_m(x)$, it follows that, by computing the zero-maps of the M members of the ensemble, one obtains an ensemble of zero-maps denoted by

$$\{Z\{G_m(w)\} = Z\{F(w)\} \cup Z\{H_m(w)\};\ \ m = 1, \ldots, M\}. \tag{6.2}$$

Since knowledge of $Z\{F(w)\}$ is essentially equivalent to knowledge of $f(x)$, it follows that the ensemble blind deconvolution problem can be re-expressed as:

> "Recover $Z\{F(w)\}$ from the ensemble $\{Z\{G_m(w)\};\ \ m = 1, \ldots, M\}$ of zero-maps specified by (6.2)."

It is advantageous to express the ensemble blind deconvolution problem in this manner because $f(x)$, and therefore $Z\{F(w)\}$, is independent of m. Thus, the zeros comprising $Z\{F(w)\}$ are contained in each of the $Z\{G_m(w)\}$. In contrast, since $h_m(x)$ is different for each m, $Z\{H_m(w)\}$ must also be different. Consequently, all the zeros that are common to all of the $Z\{G_m(w)\}$, completely characterizes

$Z\{F(w)\}$. This is succinctly described by

$$Z\{F(w)\} = Z\{G_1(w)\} \cap Z\{G_2(w)\} \cap \ldots \cap Z\{G_M(w)\}, \tag{6.3}$$

where $\cap$ is the set intersection operator. The relation embodied in (6.3) is the principle of ZAA.

Before introducing the ZAA algorithm, it is appropriate to discuss, first, properties of zeros relevant to ZAA and, second, the method in which zero-maps are displayed in this chapter.

6.1.2 Properties and Display of Zeros

In §5.3 a zero-sheet is defined as the set of locations at which the z-spectrum of an image is zero. The zero-sheets are accordingly displayed in z-space. However, when discussing ZAA it is convenient to display the zero-map in Fourier-space rather than z-space. Consequently, throughout this chapter, the notation $Z\{Q(w)\}$ denotes the set of zeros of $Q(w)$, that is, $Z\{Q(w)\}$ identifies all the points in the w-plane at which $Q(w) = 0$. Since, when $K = 1$, the zeros exist at points in a two-dimensional space, a zero-map is accurately represented as a set of geometric points lying in a plane. The location of each zero is identified by a mark placed on the plane at the appropriate place. The notation $Z\{Q(w)\}$ and the term zero-map are also invoked in this chapter to denote this graphical presentation of the set of zeros, in addition to the set of zeros itself. The intended meaning of this terminology is apparent from the context in which it appears. A general complex image and its zero-map are illustrated in Fig. 6.1(a). Here the zero locations are identified by circles, although different marks are employed in other zero-maps displayed in this chapter. This enables zeros belonging to a particular zero-map to be identified when zero-maps are superimposed.

When implementing ZAA on a digital computer, zero-maps are conveniently represented by two-dimensional arrays of pixels. These zero-maps can be subjected to the same arithmetic operations as are applied to sampled representations of two-dimensional images. In particular, zero-maps may be added together, subtracted from one another, and convolved with blurring functions.

Several useful theorems relating to complex zeros are presented by (Napier 1971, §5.3). Two important properties, from the viewpoint of ZAA, stem from the following relations:

	Image-space		Complex Fourier-space	
real	$q(x) = q^*(x)$	$\longleftrightarrow$	$Q(w) = Q^*(-w^*)$	
even	$q(x) = q(-x)$	$\longleftrightarrow$	$Q(w) = Q(-w)$	even.

The first of these relations reveals that, for situations in which the image $q(x)$ is real, its spectrum, which has been analytically continued into the w-plane, is symmetric about the v-axis. It follows, therefore, that for every zero in the right half plane there is a matching zero on the other side of the v-axis, in the left half plane. This is apparent upon inspection of the zero-map of the real image depicted in Fig. 6.1(b). The second relation reveals that, when an image is even-symmetric, its analytically continued spectrum is also even-symmetric (see Fig. 6.1(c)). By combining both properties, it is apparent that the zeros of a real, and symmetric, image are symmetric about both the origin and the imaginary axis, as shown in Fig. 6.1(d).

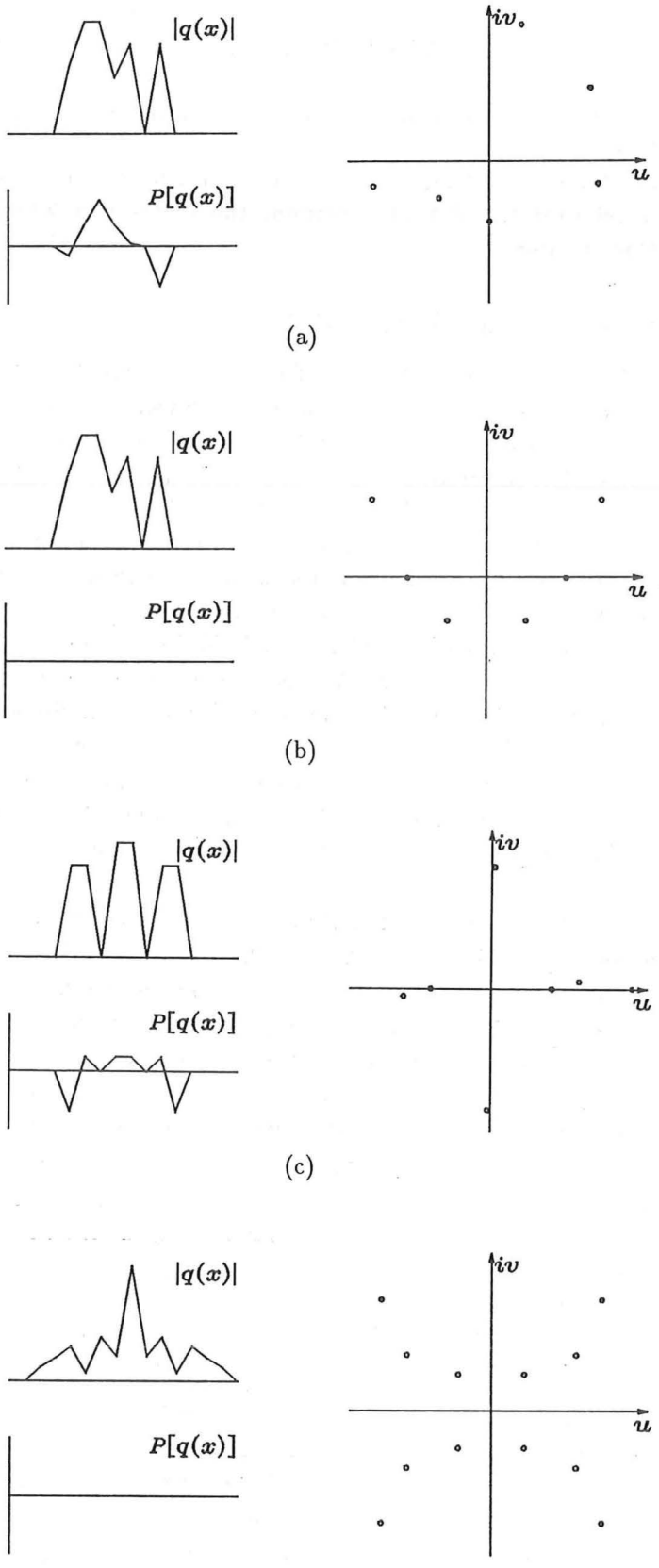
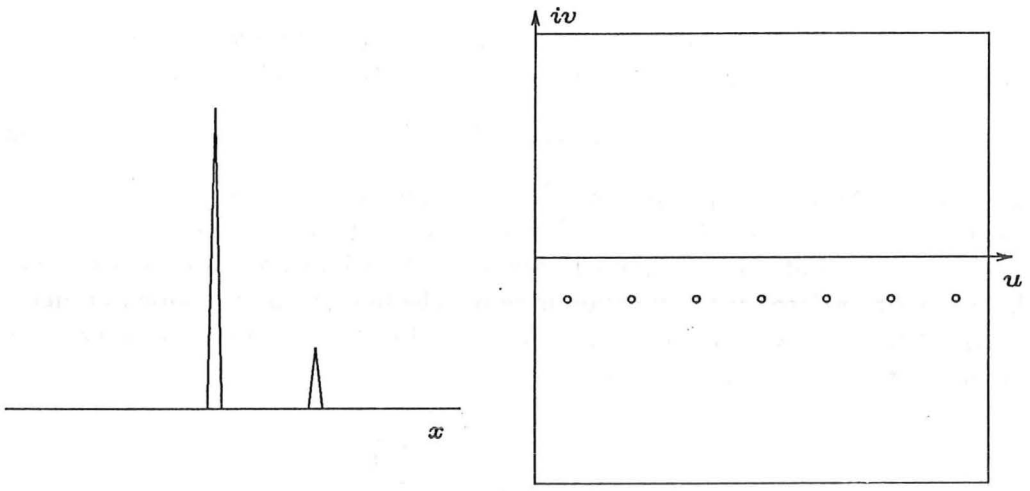


Figure 6.1: Four types of one-dimensional image and their zero-maps. The zero locations are here identified by circles: (a) $q(x)$ complex; (b) $q(x)$ real; (c) $q(x)$ complex and symmetric; (d) $q(x)$ real and symmetric.



(a) (b)

Figure 6.2: A one-dimensional binary star of 14 pixels separation and a 5:1 ratio of primary peak brightness to secondary peak brightness: (a) amplitude plot of the double star; (b) zero-map of the star's analytically continued spectrum, with the positions of the zeros indicated by circles. The u and v axes extend from 0.0 to 0.5 and -0.1 to 0.1 pixels^{-1} respectively. This scaling also applies to all subsequent zero-maps displayed in this chapter.

The examples of ZAA presented in this chapter relate to images (and to particular ways of processing them) arising in the field of optical or infrared astronomy. Since all the images of interest in this field are real (they are also positive in fact), their complex zeros are distributed in pairs reflected in the v-axis (Bates and McDonnell 1986, §13). Consequently, only the right half of the complex w-plane is displayed in the zero-maps presented in this chapter. For example, a simulated binary star which is real, and its corresponding zero-map are shown in Fig. 6.2. It is implied, when only the right half plane is shown, that an equal number of zeros exist in the left half plane. The same convention is adopted for all subsequent zero-maps displayed in this chapter.

It is important to note that, although ZAA is only invoked in this chapter with reference to positive images, it can easily be adapted to process complex images. In such cases, both half planes must be processed, because the zeros are no longer located symmetrically about the imaginary axis.

6.1.3 Computing Zeros

In practice, when processing data with a computer, the data must exist in sampled form (see §2.1.1). Such an image, $q(x)$ say, may be written as

$$q(x) = \sum_{n=0}^{N} q_n \delta(x - n\Delta x), \tag{6.4}$$

where q_n are the N samples and Δx is their spacing.

The discrete Fourier transform (2.41) of (6.4) is

$$Q(w) = \sum_{n=0}^{N} q_n e^{-\imath 2\pi n w \Delta x}. \tag{6.5}$$

When the image is pixellated it is more convenient (Fright 1984; Sinton 1986) to calculate the complex zeros of the z-transform $Q(z)$ (refer to §2.6), where

$$z = e^{-\imath 2\pi w \Delta x}. \tag{6.6}$$

An image containing $N+1$ pixels becomes a polynomial of degree N in the z-domain, with the coefficients of the polynomial given by the amplitudes of the image pixels. The zeros of this polynomial specify the locations on the complex z-plane where the z-transform of the function is zero. The fundamental theorem of algebra (Kreysig 1979, p. 653) shows that any one-dimensional polynomial can be expressed as a product of factors of the form

$$Q(z) = a_n \prod_{n=1}^{N-1} \left(1 - \frac{z}{z_n}\right), \tag{6.7}$$

where $\{z_1, \ldots, z_{N-1}\}$ is the set of zeros. When considering the zeros of the spectrum, rather than the zeros of the z-spectrum, the zeros $\{z_1, \ldots, z_{N-1}\}$ must be mapped to the complex w-plane. This is readily achieved by invoking an inversion of (6.6), namely

$$w = \frac{\ln(z)}{-i2\pi\Delta x} = \frac{-\mathcal{P}[z]}{2\pi\Delta x} + \imath\frac{\ln|z|}{2\pi\Delta x}. \tag{6.8}$$

It follows from (6.7) that a polynomial of degree N has N zeros. Consider the positive image, $q(x)$, shown in Fig. 6.2(a). The extent of this image is fifteen pixels, so that its z-transform is a polynomial of degree fourteen. Consequently, the zero-map of this image should contain seven point zeros. Inspection of Fig. 6.1(b) reveals that, indeed, this is the case.

The computation of the z-space zeros is a polynomial factorization operation. When performing this processing to generate examples for this thesis, the CPOLY polynomial zero finding routine, due to Jenkins and Traub (1972), was invoked. Recently, Hager (1987) has outlined a fast algorithm based on the FFT which can be used to find a starting point for the CPOLY algorithm, thereby accelerating the polynomial factorization procedure. This modification is not incorporated in the zero finding strategy invoked in this thesis since the speed of polynomial factorization was not important for the examples presented here. However, the routine implementation of ZAA, which would require many polynomials of large order to be factorized, would only be feasible if some such procedure could be incorporated into the software.

6.2 One-Dimensional Zero-and-Add

The ZAA ensemble blind deconvolution technique is now discussed and illustrated with examples drawn from the field of astronomical speckle imaging. The generation of the speckle images with which this section is illustrated is outlined in §6.2.1. For ease of exposition it is assumed that contamination is negligible in the initial discussion of ZAA, presented in §6.2.2. However, the harmful effects upon the ZAA algorithm of additive contamination, and truncation of speckle images, are discussed and illustrated in §6.2.3.

6.2.1 Generating One-Dimensional Speckle Images

An ensemble of 100 one-dimensional speckle psfs, $\{h_m(x); m = 1, \ldots, M\}$ was computer generated in the manner described in §4.4, and stored for subsequent processing. A typical member of the ensemble is shown in Fig. 6.3(c). The one-dimensional

object, $f(x)$, which is shown in Fig. 6.3(a), was convolved with each member of the ensemble of speckle psfs to provide an ensemble of uncontaminated speckle images, $\{s_m(x) = f(x) \odot h_m(x)\}$. A typical speckle image, generated by convolving Fig. 6.3(a) and (c), is displayed as Fig. 6.3(e). Fig. 6.3(b), (d) and (f) show the zero-maps $Z\{F(w)\}$, $Z\{H_m(w)\}$ and $Z\{S_m(w)\}$, corresponding to the images depicted in Fig. 6.3(a), (c) and (e). Since, from (5.20), the zero-map of a convolution is the union of the zero-maps of the components of the convolution, all the zeros displayed in both Fig. 6.3(b) and (d) are contained in Fig. 6.3(f). Because of the manner in which the $h_m(x)$ are generated, they all have 65 pixels. The z-transform of any such image is a polynomial of order 64, which has 64 complex zeros, half of which (i.e. 32) must lie in the right half plane. Similarly, $f(x)$ in this example is nine pixels in extent. Thus, it has four zeros that lie in the right half plane. The number of zeros shown in Fig. 6.3(f) is 36, which is the sum of the number zeros, 4 and 32, shown in Fig. 6.3(b) and (d), respectively.

Consider now the extent of $s_m(x)$. The extent of convolution theorem (2.27) states that, when images are positive, the extent of a convolution is the sum of the extents of the components of the convolution. When the images are pixellated, as all images are assumed to be in this thesis (§2.1.1), it transpires that the number of pixels in a convolution is given by the sum of the number of pixels in the components, less one (Lane 1988). Thus, since $f(x)$ and $h_m(x)$ have extents of 9 and 65 pixels respectively, it follows that $s_m(x)$ has an extent of 73 pixels. This is consistent with the 36 zeros displayed in $Z\{S_m(w)\}$ (see Fig. 6.3(f)).

The processing described in §6.2.3 also requires speckle images of an unresolvable object. Since an unresolvable object is effectively a delta function, $\delta(x)$, it follows that an uncontaminated speckle image of such an object can be written (*cf.* (4.30))

$$s_{\delta,m}(x) = \delta(x) \odot h_m(x) = h_m(x), \tag{6.9}$$

where the subscript delta signifies quantities pertaining to the unresolvable object. Consequently, an ensemble of speckle psfs $\{h_m(x)\}$, can also be regarded as being an ensemble, $\{s_{\delta,m}(x)\}$, of speckle images of an unresolvable object. Thus, a further ensemble of 100 speckle psfs was generated (see §4.4) to provide the speckle images of the unresolvable object. Note that, although all members of the new ensemble have approximately the same statistics as the $h_m(x)$, none of the actual $h_m(x)$ were taken to be a member of the ensemble $\{s_{\delta,m}(x)\}$.

6.2.2 Deconvolving Uncontaminated Speckle Images

Each zero-map, $Z\{S_m(w)\}$, is quantified by setting to unity and zero, respectively, those of its pixels that do, and do not, coincide with marks representing zeros, as explained in §6.1.2. In its basic form the ZAA algorithm consists of generating an averaged zero-map, $\bar{Z}\{S_m(w)\}$, by taking the ensemble average of the zero-maps of the M speckle images. $\bar{Z}\{S_m(w)\}$ is thus defined by

$$\bar{Z}\{S(w)\} = \frac{1}{M} \sum_{m=1}^{M} Z\{S_m(w)\}. \tag{6.10}$$

Because $f(x)$ is not dependent upon m, its zero-map is the same for all m. Consequently, when the zero-maps are averaged, the zeros comprising $Z\{F(w)\}$ reinforce in $\bar{Z}\{S_m(w)\}$. These zeros are said to be *steadfast* (Bates *et al.* 1985) and have unity amplitude in $\bar{Z}\{S_m(w)\}$. However, since the speckle psf is different in each speckle

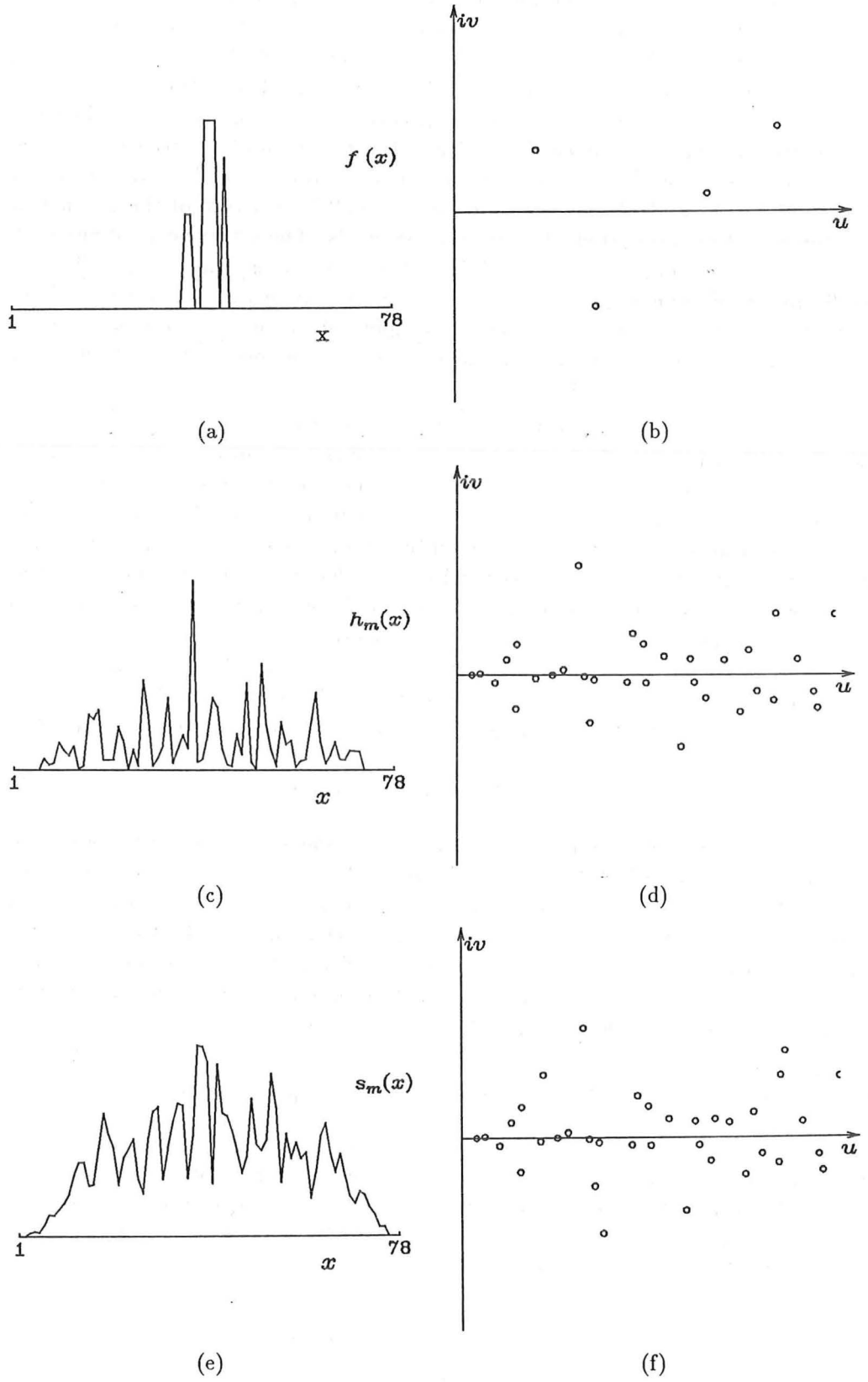


Figure 6.3: Generation of a one-dimensional speckle image and corresponding zero-maps: (a) $f(x)$; (b) its zero-map $Z\{F(w)\}$; (c) a typical speckle psf $h_m(x)$; (d) $Z\{H_m(w)\}$; (e) a typical speckle image $s_m(x)$ generated by convolving (a) and (c); (f) $Z\{S_m(w)\}$. Note that $Z\{S_m(w)\} = Z\{F(w)\} \cup Z\{H_m(w)\}$.

image, the zeros of $Z\{H_m(w)\}$ are located effectively randomly for each speckle image. Thus, except for the zeros of $H_m(w)$ that happen, by chance, to coincide with zeros of $F(w)$, $Z\{H_m(w)\}$ tends to form a background fog which has an amplitude significantly less than unity in $\bar{Z}\{S_m(w)\}$.

It is useful to define a thresholded zero-map:

$$\bar{Z}_T\{S_m(w)\} = \begin{cases} \bar{Z}\{S_m(w)\} & \text{if } \bar{Z}\{S_m(w)\} \geq \epsilon \\ 0 & \text{otherwise,} \end{cases} \tag{6.11}$$

where the subscript T indicates a thresholded zero-map, and ϵ is a positive constant that is set to unity for the contamination free case. It follows that, in the absence of contamination, $\bar{Z}_T\{S_m(w)\}$ is identical to $Z\{F(w)\}$ since these zeros are the only ones of unity amplitude in $Z\{S_m(w)\}$. Thus, the zeros remaining in $\bar{Z}_T\{S_m(w)\}$ are the zeros of $F(w)$, and it is a simple task to reconstruct $f(x)$ from these zeros (see (5.17)).

It is important to note here that, for certain seeing conditions, the $\bar{Z}\{H_m(w)\}$ exhibit apparently steadfast zeros, which are called *obdurate zeros* (Bates *et al.* 1985; Davey *et al.* 1986). The tendency is for these obdurate zeros to cluster near the imaginary axis, on both sides of the real axis. In the absence of contamination these obdurate zeros can not be as pronounced as the truly steadfast zeros, so they are eliminated in $\bar{Z}_T\{S_m(w)\}$ and therefore do not affect the fidelity of the reconstruction. However, as noted in §6.2.3, obdurate zeros can seriously degrade images reconstructed by ZAA processing unless appropriate remedial steps are taken.

6.2.3 Countering Contamination

In practice, speckle images are always contaminated so it is useful to investigate the effect of contamination upon ZAA. Contamination is usefully modelled by the term $c(x)$ added to the convolution (see (4.4)). $c(x)$ incorporates all departures of the speckle image from the convolutional model on which the imaging procedure is based. The presence of this additional term means that $s_m(x)$ is, essentially always, no longer exactly the convolution of $f(x)$ with another function (Bates *et al.* 1976, §3). Since the speckle image is compact, its analytically continued spectrum is analytic (§2.8), so an infinitesimal change in the image causes an infinitesimal change in the zero locations. Consequently, if the level of the contamination is low the zeros due to $f(\mathrm{x})$ remain in roughly the same position. However, as the level of contamination increases, the expected distance that the zeros migrate from their true positions becomes larger, and it becomes increasingly difficult to establish the true zero locations from $\bar{Z}\{S_m(w)\}$.

To counter the migration of the true zeros, it is useful to replace each zero position in each of the $Z\{S_m(w)\}$ with a blurring function, here called a *disc*. This is effectively achieved by convolving the zero-map with an appropriate two-dimensional image, which is here denoted $B(w)$. In the examples presented in this thesis,

$$B(w) = \begin{cases} 1 & \text{for } |w| < 5 \text{ pixels} \\ 0 & \text{otherwise,} \end{cases} \tag{6.12}$$

which implies the disc is chosen to have a uniform magnitude of unity and a diameter determined by the contamination level. Since the discs corresponding to any particular zero of $F(w)$ tend to overlap and form a composite disc when the $Z\{S_m(w)\}$

are superimposed, the zeros of $F(w)$ appear *quasi-steadfast* in $\bar{Z}\{S_m(w)\}$. Because of the migration of the zeros, the amplitude of the peak of the composite disc due to a particular zero of $F(w)$ can be expected to be significantly less than unity. This implies that the value of the parameter ϵ, invoked to generate $\bar{Z}_T\{S_m(w)\}$, must be reduced as the contamination level increases.

Unfortunately, blurring the zero-maps tends to enhance the effect of the obdurate zeros by merging them, thereby raising their levels in addition to raising the levels of the truly quasi-steadfast zeros. Consequently obdurate zeros often appear in $\bar{Z}_T\{S_m(w)\}$, thereby degrading the reconstruction. Thus, some means of identifying the obdurate zeros is needed. Consider now the zero-map $\bar{Z}\{S_{\delta,m}(w)\}$, generated from speckle images of an unresolvable object viewed under seeing conditions that are similar to those existing when the resolvable object was viewed. Such a zero-map can be expected to exhibit similar obdurate zeros because they depend upon the seeing and the type of processing, but not upon the object. Thus, it is possible to identify the obdurate zeros in $\bar{Z}\{S_m(w)\}$ from the observed clusters of zeros in $\bar{Z}\{S_{\delta,m}(w)\}$. This is reminescent of the need for invoking a reference object in other forms of speckle imaging (see Chapter 4), in order to calibrate for the seeing.

When compensating for the seeing-induced obdurate zeros by invoking the zero-map of an unresolvable object, it is convenient to define two new averaged zero-maps. The first, a subtracted averaged zero-map, is defined by

$$S\bar{Z}\{S_m(w), S_{\delta,m}(w)\} = (\bar{Z}\{S_m(w)\} - \bar{Z}\{S_{\delta,m}(w)\}) \odot B(w), \tag{6.13}$$

where $B(w)$ is blurring function described by (6.12). Since $\bar{Z}\{S_{\delta,m}(w)\}$ is an estimate of the seeing-induced zeros, subtracting it from $\bar{Z}\{S_m(w)\}$ effectively removes the seeing-induced zeros from $\bar{Z}\{S_m(w)\}$. Thus, since the obdurate zeros are seeing-dependent, they are effectively eliminated from $S\bar{Z}\{S_m(w), S_{\delta,m}(w)\}$. The second zero-map is the thresholded version of $S\bar{Z}\{S_m(w), S_{\delta,m}(w)\}$, which, in the manner of (6.11), is defined as

$$S\bar{Z}_T\{S_m(w), S_{\delta,m}(w)\} = \begin{cases} S\bar{Z}\{S_m(w), S_{\delta,m}(w)\} & \text{if } S\bar{Z}\{S_m(w), S_{\delta,m}(w)\} \geq \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{6.14}$$

Figs. 6.4 and 6.5 illustrate the harmful effects upon ZAA of adding contamination to the speckle images of the resolved and unresolved objects. Fig. 6.4(a)–(e) display various zero-maps generated when processing speckle images contaminated to a level of $\epsilon_{\text{dB}} = -30\text{dB}$. Obdurate zeros are manifested by the darker parts of the $\bar{Z}\{S_{\delta,m}(w)\}$ shown in Fig. 6.4(b). $\bar{Z}\{S_{\delta,m}(w)\}$ is subtracted from $\bar{Z}\{S_m(w)\}$, and the result thresholded, to generate $S\bar{Z}\{S_m(w), S_{\delta,m}(w)\}$. When this is convolved with a disk 10 pixels in diameter, the composite discs corresponding to the true steadfast zeros, become apparent. Thresholding this blurred image yields $S\bar{Z}_T\{S_m(w), S_{\delta,m}(w)\}$, as shown in Fig. 6.4(e). It is apparent that virtually all traces of the obdurate zeros have disappeared in $S\bar{Z}\{S_m(w), S_{\delta,m}(w)\}$, so that, with a suitable value of ϵ, $S\bar{Z}_T\{S_m(w), S_{\delta,m}(w)\}$ exhibits only the zeros corresponding to zeros of $F(w)$, as shown in Fig. 6.4(e). Each of the dark 'blobs' in $S\bar{Z}_T\{S_m(w), S_{\delta,m}(w)\}$ corresponds to a zero of $F(w)$ and the location of the zero is taken to be the blob's pixel of maximum magnitude. The reconstruction resulting from these four zeros, and of course the four in the left half plane, is displayed in Fig. 6.4(f). Corresponding reconstructions for levels of contamination of $\epsilon_{\text{dB}} = -40\text{dB}$ and $\epsilon_{\text{dB}} = -20\text{dB}$ are depicted in Fig. 6.4(a) and (b), respectively.

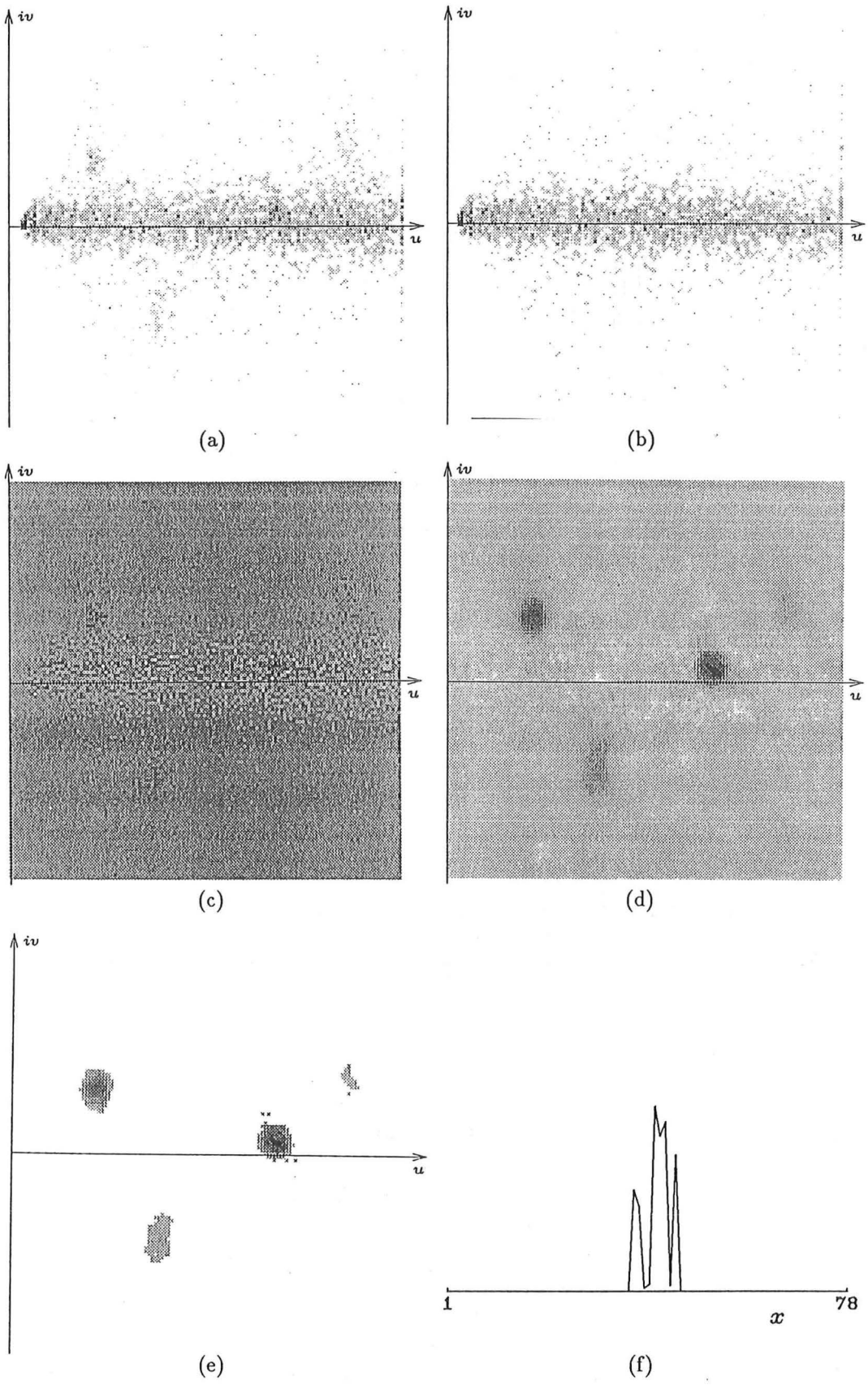


Figure 6.4: Zero-and-add when contamination to a level of $\epsilon_{\mathrm{dB}} = -30\mathrm{dB}$ is added to the speckle images: (a) $\bar{Z}\{S_m(w)\}$; (b) $\bar{Z}\{S_{\delta,m}(w)\}$; (c) $\bar{Z}\{S_m(w)\} - \bar{Z}\{S_{\delta,m}(w)\}$; (d) $S\bar{Z}\{S_m(w), S_{\delta,m}(w)\}$; (e) $S\bar{Z}_T\{S_m(w), S_{\delta,m}(w)\}$ ($\epsilon = 0.3$); (f) reconstructed image. Note that, for ease of display, the grey-scale of zero-maps (a)–(e) is inverted.

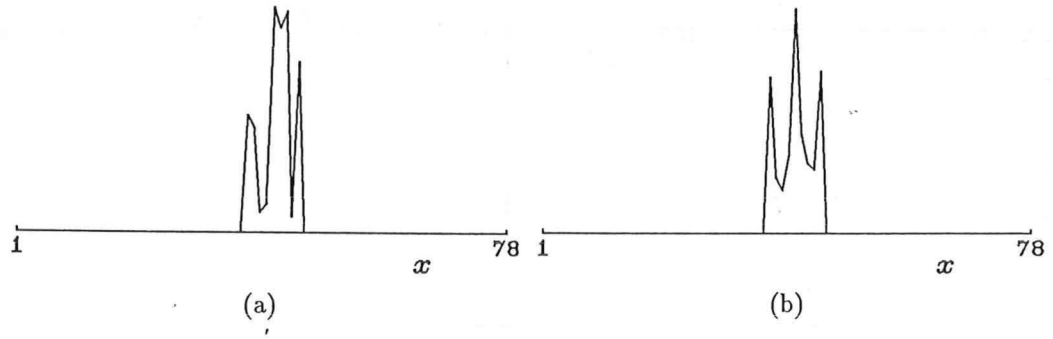


Figure 6.5: Reconstructions from zero-and-add when contamination to a level of: (a) $\epsilon_{\text{dB}} = -40\text{dB}$; (b) $\epsilon_{\text{dB}} = -20\text{dB}$, was added to the speckle images.

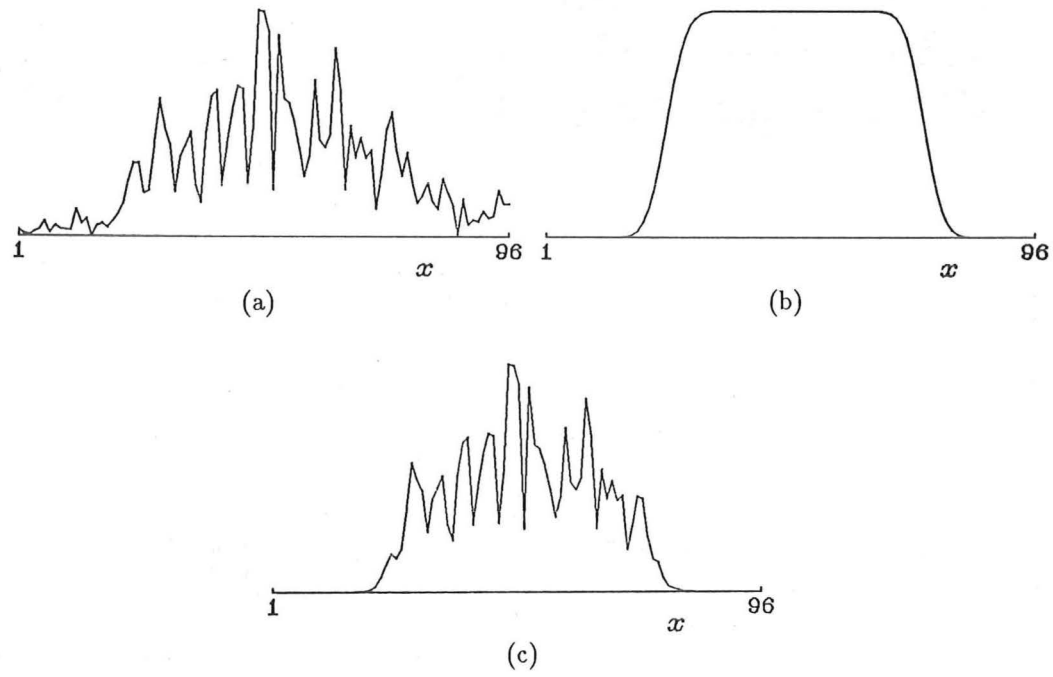


Figure 6.6: Generation of truncated one-dimensional speckle images: (a) unwindowed speckle image; (b) window function; (c) windowed speckle image.

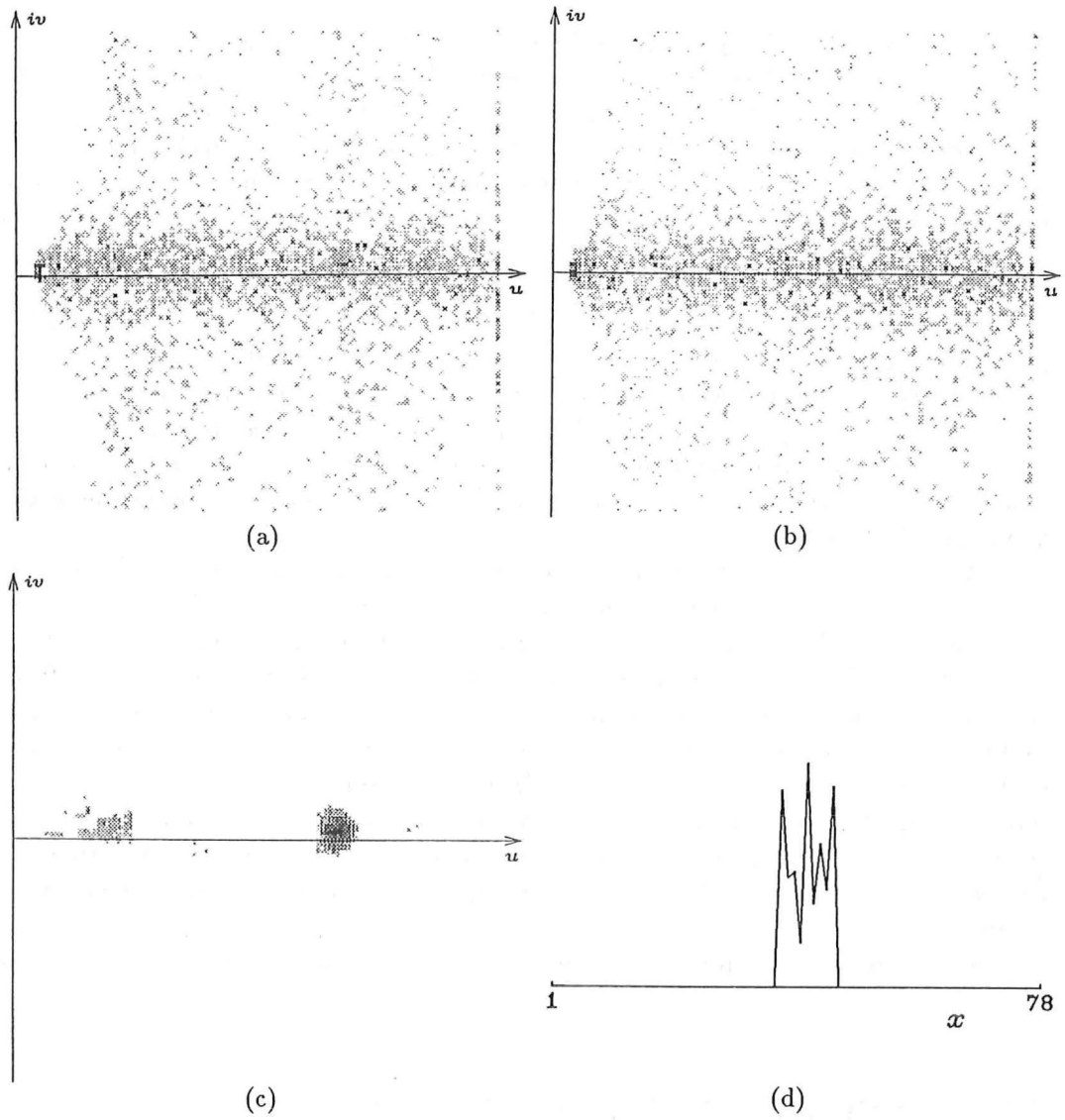


Figure 6.7: Zero-and-add processing applied to truncated speckle images: (a) $\bar{Z}\{S_m(w)\}$; (b) $\bar{Z}\{S_{\delta,m}(w)\}$; (c) $S\bar{Z}_T\{S_m(w), S_{\delta,m}(w)\}$ ($\epsilon = 0.14$); (d) reconstruction.

Although the reconstructed images shown in Fig. 6.4(f) and Fig. 6.5(a) and (b) are noticably different from the original object (Fig. 6.3(a)), the principal characteristics of that object are clearly revealed in each reconstruction. It is apparent, however, that a less faithful reconstruction results when the level of contamination is increased. For contamination levels larger than $\epsilon_{\text{dB}} = -20\text{dB}$ the composite discs were no longer distinguishable from the obdurate zeros without additional *a priori* information about the expected locations of the true zeros, or equivalently, the obdurate zeros. Consequently, no meaningful reconstruction was possible for these contamination levels.

Although the effects of moderate amounts of additive contamination do not seem overly deleterious, quite mild preprocessing (see §6.3.3) of the speckles can seriously degrade the reconstructed image. An example of such degradation is presented in Fig. 6.7. Each speckle image for this example was generated as the convolution of a speckle psf, of 256 pixels extent, convolved with the object shown in Fig. 6.3(a). A typical one of the 100 such speckle images is shown in Fig. 6.6(a). The resulting

speckle image was subsequently multiplied with the window depicted in Fig. 6.6, to reduce its extent. Fig. 6.6(c) depicts the windowed version of the speckle image Fig. 6.6(a). Zero-maps of ZAA applied to the 100 windowed speckle images are displayed in Fig. 6.7(a)–(c). Comparing Fig. 6.7(c) with Fig. 6.4(f) demonstrates how severe the effect of windowing uncontaminated speckle images can be. The reconstructed image shown in Fig. 6.7(d) is significantly inferior to the images shown in Fig. 6.4(f) and Fig. 6.5(a) and (b), even though the window shown in Fig. 6.6(b) might seem too gentle to have any significant effect.

Sinton (1986, §6.5) presents a detailed study of the affects of additive noise, photon limiting and truncation of the speckle images. He finds that a gradual deterioration in the quality of reconstruction arises when the contamination becomes more severe. Additive contamination and photon limiting tend to cause the quasi-steadfast zero to migrate towards the real axis. However, as regards windowing the speckle images, the window shape and extent both unpredictably affect the migration of the quasi-steadfast zeros.

ZAA is a novel ensemble blind deconvolution technique that can process ensembles of differently blurred one-dimensional images which have significant deviations from perfect isoplanatism. In the presence of additive contamination, the quasi-steadfast zeros in the averaged zero-map tend to become diffuse and migrate from their true positions. These effects become increasingly pronounced as the contamination becomes more severe. The migration of the zeros leads to a gradual deterioration in the reconstructed image. However, the increasing diffuseness of the quasi-steadfast zeros causes little error until such time as the zeros are so diffuse that they cannot be correctly identified in the thresholded zero-map. ZAA is based upon locating the positions of all the zeros of the image, and if a single zero is not identified severe distortion results.

Extensions to ZAA allowing it to process two-dimensional data are discussed in §6.4.

6.3 Application to Infrared Speckle Images

Astronomical speckle processing techniques (Chapter 4), which were originally intended for processing speckle images captured at visible wavelengths, have been successfully extended to infrared wavelengths (*cf.* §4.9). A new composite method (*cf.* Davey *et al.* 1989b) for processing one-dimensional infrared speckle images is presented in this section.

To be effective, most astronomical speckle imaging methods require some form of calibration for the seeing (see Chapter 4). This calibration is commonly achieved by viewing an unresolved (reference) object under seeing conditions statistically similar to those existing when the resolved object is viewed. From such measurements an estimate of the long-time average or gross effects of the atmosphere can be deduced. Unfortunately, the statistics describing the atmosphere are never exactly stationary (Roddier 1981), so the calibration can be expected to be significantly in error for any particular speckle image. Observing a reference object also significantly increases the telescope observation time and is therefore expensive in practice. Thus, techniques that allow atmospheric effects to be removed without the need for a reference star would be desirable. The new technique described in this section achieves such 'self-calibration'. This technique is composite in that six distinct processing techniques are combined to extract the true image from a set of infrared speckle images. The steps in the technique are:

(i) estimate the instantaneous seeing quality existing at the time each speckle image was recorded,

(ii) bin (see §6.3.1) each speckle image according to the estimated seeing quality,

(iii) apply shift-and-add (SAA) to the ensemble of images contained in each bin,

(iv) invoke edge-extension (see §6.3.3) to reduce the effect of contamination,

(v) apply an adaptation of the ZAA technique to identify the zeros due to the object in the edge-extended SAA images, and

(vi) from these zeros, reconstruct an estimate of the true object.

Approaches to retrieving the brightness ratio and separation of the binary star from the zero clusters are outlined. The first is a statistical technique due to Jeffreys (1980) which, in its basic form, takes no account of deficiencies in the method employed to record the one-dimensional images.

In practice, a significant source of contamination of one-dimensional speckle images is the nonlinear scanning motion of the secondary mirror of the telescope (Leinert and Dyck 1983; McCarthy *et al.* 1987). The composite method presented in this section is capable of compensating for nonlinear distortions in the scanning mechanism even though the detailed form of this quasi-random motion is unknown.

The method is illustrated processing infrared speckle images of the astrometric binary Ross 614 AB (Christou *et al.* 1987a), an object of considerable interest to astrophysicists (Probst 1977).

6.3.1 Data Recording and Binning

When processing ensembles of images, in situations in which each member of the ensemble is of slightly different 'quality', it is often useful to *bin* the images according to this quality. The actual criteria for deciding the quality of a particular image varies from situation to situation. Several *bins*, one for each foreseeable range of quality, are defined, the quality of each image is assessed and each image is 'deposited in' the bin that corresponds to its quality. Once every image has been binned, the 'subensemble' of images contained in each bin is processed separately from the images contained in the remaining bins. Thus, if one assigns the images to, say, L bins, then L different reconstructions result from processing the subensembles.

The motivation for binning images is that the reconstructions obtained from processing the images assigned to the 'higher quality' bins are of superior quality than the the single reconstruction generated from processing the original ensemble of images. Of course, the reconstructions from the 'lower quality' bins are less faithful, but, provided the ensemble is sufficiently large, the low quality reconstructions are often best disregarded since they contain little information anyway. An added advantage of binning the data, is that one obtains several different reconstructions because, if the assignment of images to the bins is meaningful, no two reconstructions are the same. These multiple reconstructions can be exploited by particular processing algorithms, and indeed, are exploited in the composite scheme presented here.

When applying speckle imaging techniques, several researchers have noted that accurate assessment of the instantaneous state of the atmosphere at the time when each speckle image is recorded provides an estimate of the quality of *seeing* (§4.1). Since the quality of a speckle image is essentially determined by the seeing, this

l	No. of images
1	9
2	25
3	102

Table 6.1: The numbers of speckle images of Ross 614 AB assigned to the bins corresponding to the three best conditions of seeing (i.e. $l = 1, 2, 3$).

estimate is a useful measure of the quality of a speckle image. By binning each speckle image according to its quality and applying a speckle imaging algorithm to the images contained in the bins corresponding to 'high quality', the fidelity of the reconstruction of the object is improved (*cf.* Aime *et al.* 1978; von der Lühe 1984). Such preliminary processing has also been shown to be useful at infrared wavelengths (Mariotti *et al.* 1983; Christou *et al.* 1987a).

McCarthy and Cobb (1986) show that the image sharpness parameter $S1$, defined by Muller and Buffington (1974), is related to the extent (§2.1.4) of a speckle image, and therefore also to the quality of the seeing existing at the instant when the speckle image is recorded. This parameter is defined as

$$S1 = \frac{\left(\int_{-\infty}^{\infty} s_m(x)\,\mathrm{d}x\right)^2}{\int_{-\infty}^{\infty} s_m^2(x)\,\mathrm{d}x} \tag{6.15}$$

where x_m is the point of greatest amplitude in the m^{th} speckle image.

The average quality of the seeing can be usefully characterized by a single parameter, the Fried seeing parameter, r_0 (§4.1). When calculating the quality of the seeing for a single speckle image it is useful to introduce a second parameter r_e, which is an instantaneous estimate of r_0. Statistical fluctuations in the seeing imply that, in general, r_e differs from r_0. Christou *et al.* (1987a) compute r_e by constructing a least-squares fit to the low-frequency component of the normal statistical model (Roddier 1981) of the turbulent atmosphere. They note that there always seems to be close correlation between $S1$ and r_e. Because $S1$ is significantly easier to compute than r_e, it is appropriate to take the inverse of $S1$ to be an estimate of the instantaneous value of r_e.

Christou *et al.* (1987a) demonstrated the usefulness of the parameter $S1$ by recording speckle images of several infrared sources and binning them according to the value of $S1$ computed for each of the images. Of particular interest here are the speckle images they recorded of the astrometric binary Ross 614 AB, because it is this data with which the processing described in this section is illustrated. Nine blocks, each of 512 speckle images, were recorded in the standard photometric K band ($\lambda_0 = 2.2\mu$m, $\Delta\lambda = 0.6\mu$m) on the night of October 29, 1985, on the Steward Observatory 90 inch telescope. They were recorded over a period of approximately 50 minutes with a North-South scan direction and an image scale of 76marcsec per pixel. Four of these speckle images are depicted in Fig. 6.8. In addition, blocks of 512 speckle images of the unresolved star SAO 133312, which is near Ross 614 AB, were recorded alternately with those of Ross 614 AB. Four of the speckle images of SAO 133312 are depicted in Fig. 4.14.

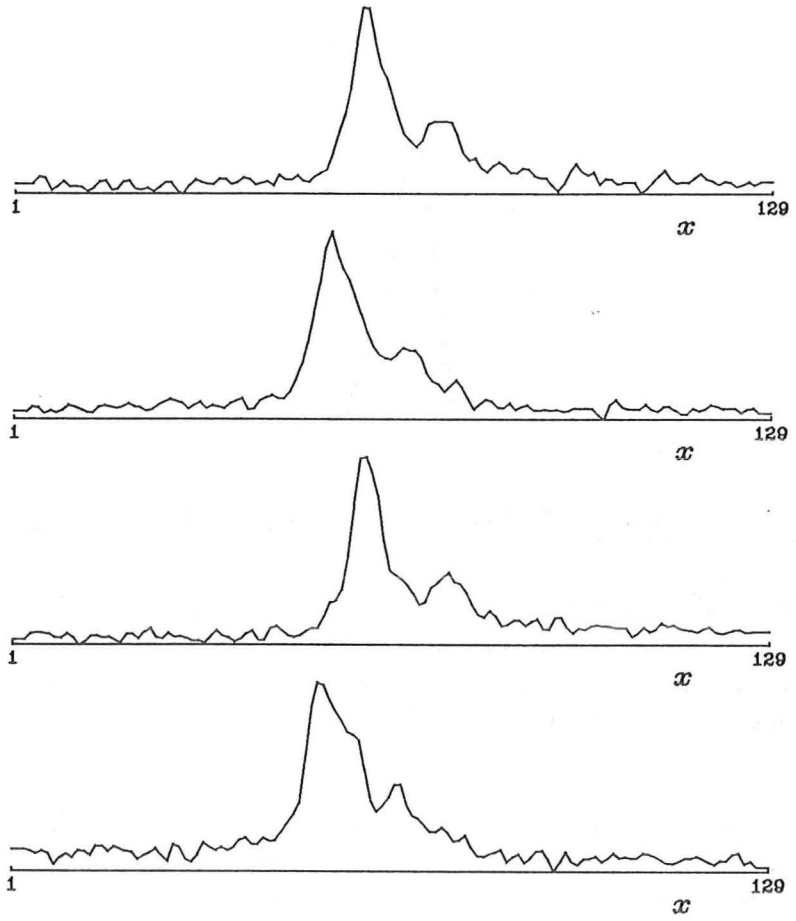


Figure 6.8: One-dimensional infrared speckle images of the astrometric binary star Ross 614 AB. The method presented in this section is illustrated by processing speckle images of this star.

6.3.2 Shift-and-Add of Binned Data

When the data are assigned to L bins according to image quality, the SAA images so produced are conveniently written as (*cf.* (4.28))

$$f_{\rm sa}{}^{(l)}(x) = f(x) \odot h_{\rm sa}{}^{(l)}(x) + \tilde{c}^{(l)}(x);\ l = 1, \ldots, L, \tag{6.16}$$

where the superscript (l) indicates the $l^{\rm th}$ bin of data and $l = 1$ and $l = L$ correspond to the best and worst seeing conditions respectively. When basic SAA (§4.8.1) was separately applied to each of the bins, the image produced from the bin corresponding to the best seeing (i.e. r_e being greatest) was of superior quality to the SAA image generated by applying the algorithm to all of the recorded speckle images. Christou *et al.* (1987a) used $L = 18$ for binning the Ross 614 AB data. The number of speckle images assigned to bins $l = 1, 2, 3$ are listed in Table 6.1.

SAA is performed separately for each of the bins, thereby generating several images comprising the convolution of the same true object, $f(x)$, with a SAA psf, $h_{\rm sa}{}^{(l)}(x)$, which differs for each l. The SAA images corresponding to the three best seeing conditions are shown in Fig. 6.9. The image corresponding to the best seeing $(l = 1)$ $f_{\rm sa}{}^{(1)}(x)$ reveals the two components of the binary star clearly resolved, even though the number of images assigned to the $l = 1$ bin is comparatively small. The two components are also apparent in $f_{\rm sa}{}^{(2)}(x)$ and $f_{\rm sa}{}^{(3)}(x)$, but in these images the

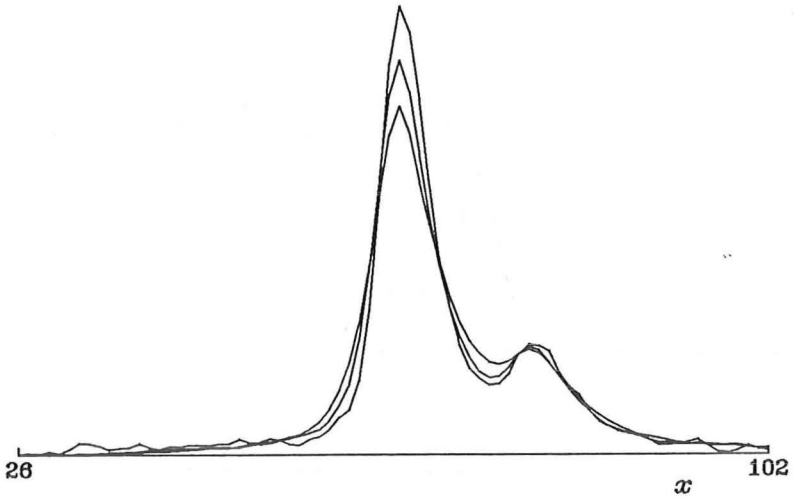


Figure 6.9: SAA images of the binary star Ross 614 AB generated from the images assigned to bins $l = 1, 2, 3$ (as published in Christou *et al.* 1987a). As l increases, the peak of the resulting SAA image decreases and the secondary peak becomes less pronounced. The extent of the figure in the x-direction is 64 pixels (4.9arcsec).

secondary peak is less pronounced than that in ${f_{\rm sa}}^{(1)}(x)$. The quality of ${f_{\rm sa}}^{(l)}(x)$, for $l > 3$, deteriorates rapidly with increasing l. While the smaller component can be recognized with hindsight in some of these SAA images, its apparent amplitude is significantly different for the SAA images corresponding to $l > 3$. This is why it is only the data in the three 'best' bins that are permitted to contribute to the results reported here.

The conventional method of obtaining an estimate of $f(x)$ from $f_{\rm sa}(x)$ is to Wiener filter (see §3.1.2) $f_{\rm sa}(x)$ with $\hat{h}_{\rm sa}(x)$, an estimate of $h_{\rm sa}(x)$. The Wiener filter is a modification of the simple inverse filter (§3.1.1), in which an additional term, $\Phi(u)$, is incorporated to ameliorate the harmful effects of contamination on the filtering operation. $\Phi(u)$ is optimally the noise-to-signal ratio for each spatial frequency u. Since the detailed dependency of this ratio upon u is not usually known, it is convenient to replace $\Phi(u)$ with a constant Φ, which is termed the *Wiener filter constant.* In practice, the optimum value of Φ is usually unknown *a priori*. In such a situation the Wiener filter can be applied several times, each time with a different Φ, with the reconstruction that is 'best' according to some criteria being chosen as the best estimate of the object. This is why Table 6.2 lists brightness ratios and separations obtained by applying a Wiener filter with several different values of the Wiener filter constant Φ.

$\hat{h}_{\rm sa}(x)$ is most straightforwardly generated by performing SAA on speckle images of an unresolved object, viewed under statistically similar seeing conditions to those existing when the resolved object was viewed (see §4.8.1). Thus, for the speckle images of Ross 614 AB with which this this section is illustrated, the speckle images of the unresolvable star SAO 133312 provide such reference information. Consequently, the same processing that was applied to the Ross 614 AB speckle images was also applied to those of SAO 133312. The SAA images of SAO 133312 corresponding to the three best seeing conditions are shown in Fig. 6.10. Each of these images was used to Wiener filter (*cf.* (3.8)) its corresponding $f_{\rm sa}(x)$ to obtain an estimate of Ross 614 AB. The separations and brightness ratios found for the two components

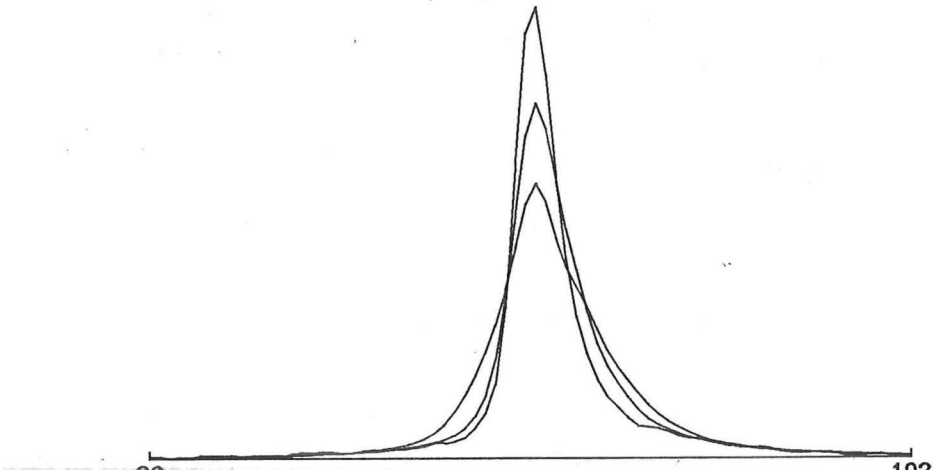


Figure 6.10: SAA images of the unresolved star SAO 133312 from the bins corresponding to the 3 best conditions of seeing. Several of the speckle images of SAO 133312 are shown in Fig. 4.14.

γ	$l=1$		$l=2$		$l=3$	
	ρarcsec	r	ρarcsec	r	ρarcsec	r
0.0003	1.04	6.31	1.02	5.24	1.03	4.97
0.001	1.04	6.16	1.02	5.11	1.03	4.70
0.003	1.03	6.04	1.02	5.10	1.03	4.39
0.01	1.03	5.86	1.02	4.95	1.06	4.37

Table 6.2: Separation (ρ) and brightness ratio (r) obtained by Wiener filtering the 'binned' SAA images of Ross 614 AB with corresponding binned SAA images of the unresolved star SAO 133312. The values shown here were generated by least-squares fitting a quadratic function to the five pixels at the peaks corresponding to the components of the star. The filter constant was taken to be $\Phi = \gamma \max_u |\hat{H}_{\rm sa}(u)|^2$. Note the large variation, with l and γ, in both the separation and brightness ratio.

of Ross 614 AB calculated for several values for the Wiener filter constant, Φ, are listed in Table 6.2.

The results presented in Table 6.2 exhibit appreciable dependence upon both l and Φ. The dependence upon l arises from the inexact matching of the seeing quality of the bins of speckle images of the resolved star with the corresponding bins for the reference star. The dependence upon Φ is a result of (3.8) not being a pure inverse filter. The Φ term adds a bias to the deconvolution and, as Φ becomes larger, so does the bias, and hence the error in the result. Since it is virtually impossible to quantitatively estimate contamination levels with confidence, one cannot be certain of the optimum value for Φ. Because of the dependence of the values of separation and brightness ratio upon l and Φ, one is unable to deduce their 'best' values from the Wiener filter operation. There is the further point that $h_{\rm sa}(x)$ is object dependent, in that its form is determined both by the seeing conditions and the form of $f(x)$ (see §4.8.2). This implies that, even if the optimum value for Φ were known,

and if the seeing of the bins could be matched exactly, it might not be of much help, because significant differences between $h_{\rm sa}(x)$ and $\hat{h}_{\rm sa}(x)$ are to be expected for certain classes of object. The processing described in §§6.3.3 and 6.3.4 below overcomes these difficulties.

6.3.3 Improving Shift-and-Add Images with Zero-and-Add

A binary star, whose individually unresolvable components are separated by Δl pixels and have a brightness ratio r, can be represented by (*cf.* (4.5))

$$d(x) = r\delta(x) + \delta(x - \Delta l) \tag{6.17}$$

where $\delta(x)$ is the Dirac delta function, and $r = 1/R$. Note that, for convenience, the term brightness ratio is invoked in this chapter to denote the reciprocal of the quantity it is employed to denote in Chapter 4. The zeros of $D(w)$, which is the Fourier transform of $d(x)$, are found by replacing $q(x)$ in (6.4) with $d(x)$, as defined by (6.17), computing $D(w)$ (*cf.* (6.5)), and setting it to zero. These zeros are located at

$$w = \frac{2n-1}{2\Delta l} + \imath\frac{\ln(r)}{2\pi\Delta l};\ n = -\infty, \ldots, \infty. \tag{6.18}$$

A representation of a two component star, for which $\Delta l = 14$ and $r = 5.0$, is shown in Fig. 6.2(a), and its zero-map is depicted in Fig. 6.2(b). Inspection of the zero-map reveals that the zeros of a binary are equally spaced along a line parallel to the real axis. It follows from (6.18) that the spacing between zeros is dependent only upon the separation, whereas the distance of the zeros from the real axis is dependent upon both Δl and r.

The ZAA principle (§6.1.1) states that, when zero-maps of N differently blurred versions $b^{(n)}(x)$ of $d(x)$, each formed by convolving $d(x)$ with a psf $h_{\rm sa}{}^{(n)}(x)$, are superimposed, the zeros of $D(w)$ reinforce while the zeros of the $H_{\rm sa}{}^{(n)}(w)$ are quasi-randomly scattered over the complex w plane. The zeros of $D(w)$ can thus be identified by inspection. Since each of the binned SAA images of Ross 614 AB is the convolution of the object with a psf that is different for each bin, the zero-maps of the SAA images of the binary Ross 614 AB would be expected to exhibit clusters of zeros similar in location to the zeros shown in Fig. 6.2(b).

Before computing the zeros of the SAA images it is useful to preprocess these images. The preprocessing technique that has proven most useful in this application is edge-extension (Bates and McDonnell 1986, §15). The philosophy behind preprocessing in general, and edge-extension in particular, is briefly summarized in the following paragraph.

In many situations a recorded image contains useful information in only a portion of its pixels, with the rest being dominated by contamination. In such a situation it is sensible to preprocess the image by extracting out the pixels containing 'good' data for further processing. It is important, when extracting out the portion of an image that contains good data, that adequate precautions are taken to minimize further corruption of the data. For instance, if the image is simply truncated, discontinuities, or *glitches*, may be introduced. Glitches cause undesirable artefacts, known as *ringing* in the spectrum, with the consequence that subsequent processing of the spectrum is degraded. In contrast, edge-extension is designed to minimize the glitches in an image while still reducing the extent of an image. Edge-extension involves replacing the obviously contaminated parts of the image with 'smooth' extensions which fall to zero appropriately. The extensions are chosen to maintain continuity of amplitudes and first derivatives of the image. Edge-extension is not an attempt at any sort of

'super-resolution'. It should be mentioned in passing that windowing an image is perhaps the most widely applied preprocessing technique. Here, an image is multiplied with a window function having smoothly tapering edges, similar to that depicted in Fig. 6.6(b). In many situations windowing is effective, although it often has the disadvantage of destroying useful data in the image. For the processing reported here, both windowing and edge-extension were applied to reduce the extent of the image. Of the two techniques it was found that edge-extension produced significantly superior results.

Although each speckle image of Ross 614 AB was recorded with 128 pixels, the useful data contained in each image exist only near the centre of the image. The same is true for the SAA images, whose outer parts are dominated by contamination. Thus, edge-extension is invoked to reduce the deleterious effect of the contamination on the ZAA process. Edge-extension is applied by first truncating each SAA image where contamination has obviously affected the data. Then tails of a gaussian distribution are butted up to each side of the truncated data, maintaining continuity of the image and its first derivative with respect to x. Edge-extension has the added advantage of reducing the extent of the image. This is significant because the CPOLY polynomial factorizing routine is more stable for polynomials of lower order. The edge-extended SAA image of Ross 614 AB corresponding to $l = 2$ is depicted in Fig. 6.11(b). Comparison with the original image shown in Fig. 6.11(a), which is taken from Fig. 6.9, illustrates the contamination suppression and extent reduction achieved by the edge-extension process.

The zero-maps of the three edge-extended SAA images are shown superimposed in Fig. 6.12. Also plotted on Fig. 6.12 are the zero-maps corresponding to the three edge-extended SAA images of the reference star SAO 133312.

It is apparent from inspection of the zeros in Fig. 6.12 that there are four clusters of three zeros. In each cluster there is one zero from each of the SAA images of Ross 614 AB. These clusters can be taken to correspond to the zeros of $F(w)$ since the zeros of $H_{\rm sa}{}^{(l)}(w)$ vary with l and spread over the w-plane. The four clusters of zeros appear to correspond to the zeros of a binary star (cf. (6.18) and Fig. 6.2) since they nearly lie on a line parallel to the real axis. Departures of the zeros in the clusters from a single line parallel to the u-axis are significant because they are a direct consequence of the non-linear scanning motion (as explained in §6.3.4). None of the zeros corresponding to SAO 133312 show any tendency to form similar clusters. Since SAO 133312 is an unresolvable star it should only give rise to zeros spread widely over the w-plane, like the zeros of $H_{\rm sa}{}^{(l)}(w)$.

Note that, for large values of u, the zeros that would be expected for a binary star do not appear as clusters in Fig. 6.12. However these zeros correspond to high spatial frequencies, which tend to be far more severely affected by contamination than low spatial frequencies. Also, the diffraction limit of a 2.28m (90 inch) telescope operating at 2.2μm is approximately 200marcsec (*cf.* (4.1)). This limit is almost three times the 76marcsec sampling rate used for recording the data. Therefore, the information contained at high spatial frequencies is beyond the diffraction limit of the telescope. For both reasons the locations of the low spatial frequency zeros are expected to be more faithful than those at high frequencies.

The statistical nature of the data, arising from both the atmospheric distortion and the presence of contamination, can be dealt with by invoking a least-squares analysis method of Jeffreys (1980) (Davey *et al.* 1988a). The postulation here is that the effect of contamination on the recorded image is to move the zero positions in a quasi-random manner. Therefore, a least-squares analysis applied to the zeros

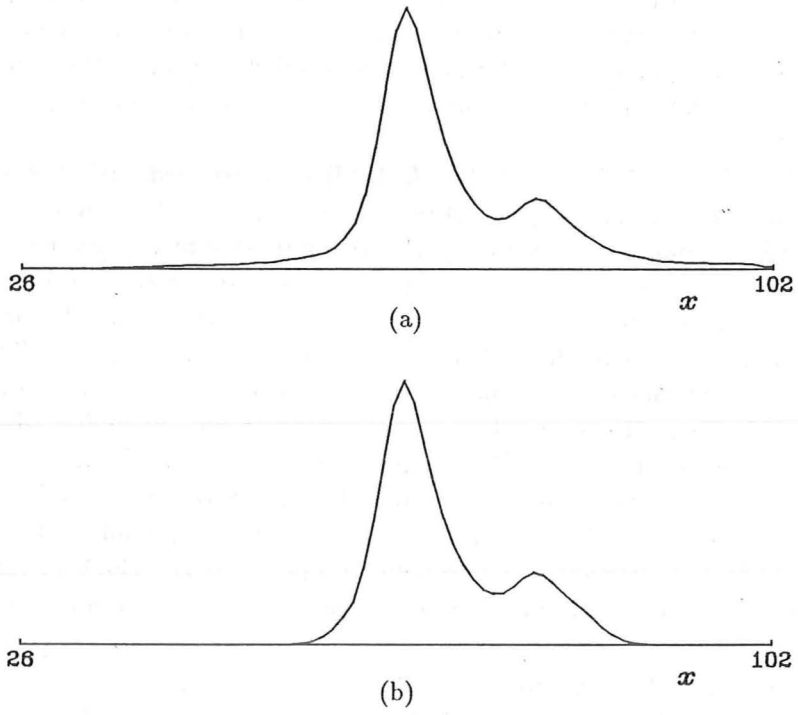


Figure 6.11: Edge-extension of $f_{sa}{}^{(2)}(x)$ (i.e. the SAA image of Ross 614 AB produced from the second bin of data) which is shown in Fig. 6.9: (a) $f_{sa}{}^{(2)}(x)$; (b) truncated and edge-extended version of $f_{sa}{}^{(2)}(x)$. The edge-extension process involved truncating the SAA image and fitting sections of Gaussian functions to the truncated regions, ensuring that the image and its first derivative at the junction are continuous. Note the reduction in the extent of the image and also the elimination of the most severely contaminated parts of the SAA image.

$\frac{1}{\sigma}$				ρ arcsec	R	σ_ρ arcsec	σ_r
$n=1$	$n=2$	$n=3$	$n=4$				
$\frac{1}{0.002}$	$\frac{1}{0.004}$	$\frac{1}{0.006}$	$\frac{1}{0.008}$	1.04	4.3	0.01	0.4

Table 6.3: Results of applying a least-squares analysis (after Jeffreys (1980) to the zeros comprising the four clusters identified in Fig. 6.12. The values $1/\sigma$ are the weighting given to the various clusters. n is the 'cluster number' which increases with increasing distance from the imaginary axis (i.e. $n = 1$ is the cluster closest the origin). σ_ρ and σ_r are the standard deviation of ρ and r respectively.

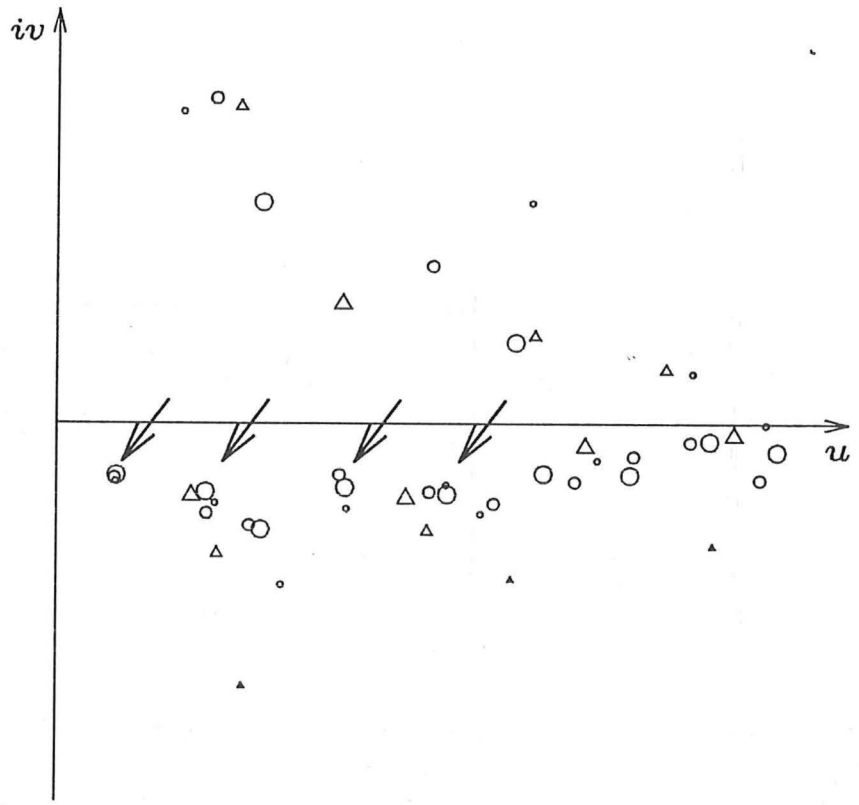


Figure 6.12: Zero-maps of the three edge-extended versions of the SAA images of Ross 614 AB (drawn as open circles) and the three of SAO 133312 shown in Fig. 6.10 (drawn as open triangles). The zero positions represented by markers (i.e. circles or triangles) of the same size and shape correspond to the same SAA image, and increasing marker size corresponds to decreasing seeing quality (i.e. increasing l). Four clusters, each identified by an arrow and each containing one zero from all three of the SAA images of Ross 614 AB, are located along a line below the real axis. It is the zeros comprising these four clusters that are invoked in subsequent processing.

found in each cluster provides a useful estimate of the true zero position as well as a measure of the error. This analysis is based upon the assumption that the star being observed is a binary, so the zeros lie on a line parallel to the real axis. The binary is completely described by the separation of the zeros, and the distance of the zeros from the real axis. Thus, it is these two parameters that are required to be reconstructed from the clusters by invoking the least-squares analysis. Furthermore, since the location of the zeros at low frequencies are expected to be more faithful than those at high frequencies, the closer a cluster is to the origin, the more weight it is given in the least-squares analysis. The actual weightings are listed in Table 6.3. Further details of this analysis are provided by Davey *et al.* (1988a).

The estimates of separation and brightness ratio, and their standard deviations, generated by applying the statistical analysis procedure to the zeros constituting the four clusters marked by arrows in Fig. 6.12, are presented in Table 6.3. Note that the separation and its standard deviation are given in angular measure (i.e. ρ) rather than pixels (Δl). Although the analysis gives the standard deviation of the separation as 0.01arcsec, it is unrealistic to specify an accuracy less than the 0.076arcsec image scale used in the sampling. Thus, the estimates for the separation and brightness ratio of Ross 614 AB from this least-squares analysis are 1.04 ± 0.08arcsec and 4.3 ± 0.4.

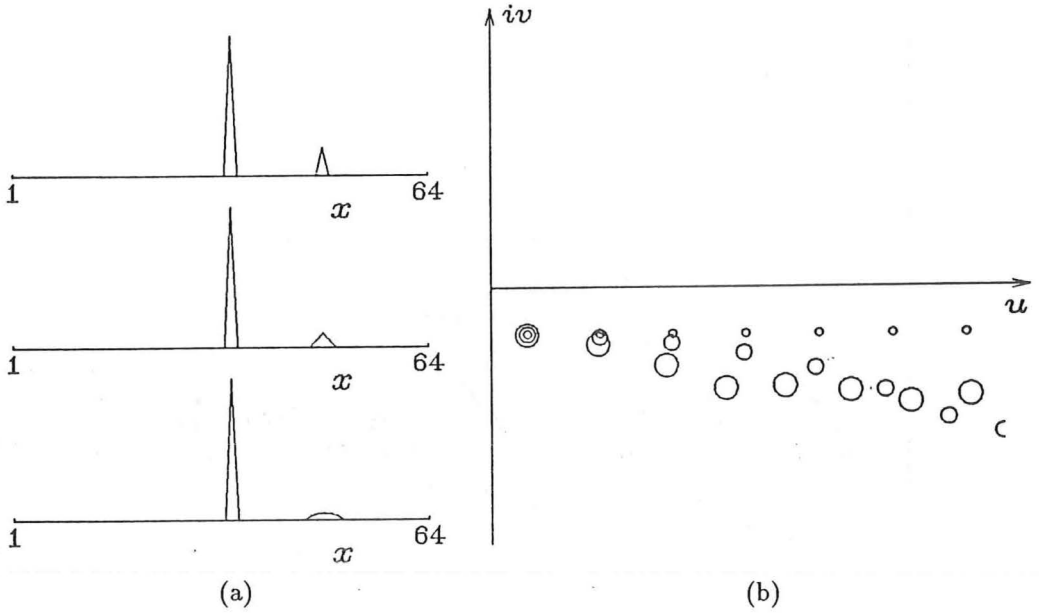


Figure 6.13: Effect upon the zero-map of the spreading of the secondary peak of a binary star: (a) Three simulations of a binary star with varying amounts of secondary peak spreading. The sum of the pixels comprising the secondary peak is identical in each of these images; (b) zero-maps of the three images shown in (a). The zero positions are depicted by open circles, with the increasing size of the circle representing increasing spreading of the secondary peak.

6.3.4 Taking Explicit Account of Nonlinear Scanning Motion

Because of the nonlinear nature of the motion of the secondary mirror (McCarthy *et al.* 1987), the scale of the x-axis of each individual speckle image is slightly distorted in some unknown manner. This means that the measured separation of the components varies from speckle image to speckle image. The mean separation is of course expected to equal the true separation.

It follows that, in this situation, the RHS of (4.28) is not a faithful representation of a SAA image. Instead, $f(x)$ should be replaced by a distorted version, $\tilde{f}(x)$, of the true image, which accounts for the variable separation. In the case of a binary, for which $f(x)$ comprises two separated pixels (as in Fig. 6.2(a)), the pixel representing the fainter (secondary) component of the binary should be spread out, in a plot of $\tilde{f}(x)$, into a number of contiguous pixels, as indicated by the simulated images shown in Fig. 6.13(a). Note that the sum of the collection of pixels, comprising the fainter component of the binary in each of the three projections shown in Fig. 6.13(a), is the same and also equal to the smaller pixel in Fig. 6.2(a).

Consider the zero-maps, depicted in Fig. 6.13(b), of the images that have been distorted by secondary peak spreading, which are shown in Fig. 6.13(a). There are two noteworthy points about the distribution of zeros in these zero-maps. Firstly, the position of the zero closest to the origin is almost identical for each of the four zero-maps shown in Fig. 6.13(b). Secondly, in the presence of secondary peak spreading the complex zeros still lie on a straight line, although the line is no longer parallel to the real axis.

Consider the four clusters of zeros identified by arrows in the zero-maps of the edge-extended versions of Ross 614 AB (Fig. 6.12). Of these clusters the position of

the three zeros comprising the cluster closest to the origin (i.e. $n = 1$) are virtually identical. The other three zero clusters seem to lie on a straight line, reminescent of those zeros shown in Fig. 6.13(b). This evidence supports the assertion that the secondary peak is spread in the manner depicted in Fig. 6.13(a). Thus, it is possible to compensate for the nonlinearity of the scanning by least-squares fitting a straight line to the zeros comprising the four clusters. The line, and the zeros, are rotated until the line is parallel to the real axis. In this manner the distortion in the location of the zeros, induced by the non-linear scanning motion, is eliminated. When fitting the line to the twelve zeros comprising the four clusters, the zeros in each cluster were weighted with the uncertainties $(1/\sigma)$ listed in Table 6.3. The line was then rotated to be parallel to the real axis, thereby eliminating the distortion due to the peak spreading. Since, in Fig. 6.13(b), the zeros comprising the cluster $n = 1$ are essentially invariant to small amounts of distortion, it is sensible to choose the point about which to rotate the fitted line according to this cluster. Thus, the line was rotated about a point defined by the intersection of the line itself with a line parallel to the real axis and passing through the average real part of the zeros comprising the first cluster of zeros. The least-squares technique of §6.3.3 was then reapplied to the new, rotated, zero positions. This produced estimates of the brightness ratio and separation of Ross 614 AB of 3.9 ± 0.4 and 1.04 ± 0.01arcsec, respectively. Note that although the separation is identical to that obtained from the unrotated data, the reconstructed brightness ratios differ, although by less than the estimated error.

6.3.5 Summary

Least-squares analysis applied to the unrotated zero-positions yields a separation and brightness ratio at infrared of the two components of Ross 614 AB of 1.04 ± 0.08arcsec and 4.3 ± 0.4. These values compare very favourably with the values 1.04arcsec and 4.37 respectively, computed by applying triple correlation processing (§4.7) to all the speckle images of the same data set (Christou *et al.* 1987b). The results obtained after compensation for the distortion introduced by the non-linear scanning mechanism are 1.04 ± 0.08arcsec and 3.9 ± 0.4. It is important to note, first, that Christou *et al.* do not take complete account of the non-linear scanning mechanism and, second, that the technique presented here is significantly less computationally demanding than triple correlation processing.

The new composite scheme for processing the one-dimensional infrared speckle images that is presented in this section has the following advantages:

(i) It avoids the need for observing a reference star under 'statistically similar' seeing conditions. This advantage is significant because it is difficult in practice to make accurate quantitative comparisons of different seeing conditions.

(ii) It overcomes the problem inherent in basic SAA (§4.8.1) that the composite psf is not determined solely by the seeing conditions, i.e. $h_{\mathrm{sa}}(x)$ depends both upon these conditions and the detailed form of $f(x)$.

(iii) It allows the quality of individual speckle images to be assessed, so that the images can be graded and stored in separate 'bins'. The important point here is that the grading does not have to be done with any great precision. The reason for this is that the sum of any number of differently blurred versions of a particular image (each blurred version being shifted by an arbitrary amount) can be expressed as the convolution of this image with another psf (i.e. a psf different from all of those causing the blurring in the aforesaid blurred versions).

The only crucial requirement on the grading is that the psfs implicit in the SAA images obtained from the different bins should themselves be significantly different. This requirement is met by ensuring the statistical characters of the speckle images stored in different bins are as different as practicable.

(iv) It preserves the computational simplicity of basic SAA for the majority of the processing.

(v) It allows straightforward compensation for nonlinear scanning in the absence of detailed knowledge of the scanning motion.

This technique has important practical implications for speckle processing because of its simplicity. Basic SAA is a simple technique and it can easily be performed on a microprocessor. The parameter $S1$ which provides an estimate of the seeing quality for binning the data is also computationally undemanding. By comparison with other extensions to SAA (Minard *et al.* 1985) and other speckle processing methods (Bates 1982b; Freeman *et al.* 1987) which require sophisticated processing of each speckle image, this technique limits intensive computation (e.g. blind deconvolution by ZAA) to only a few SAA images, each of which is formed from the speckle images collected in a single bin.

The processing scheme is most useful when a valid *a priori* model of the object being viewed (e.g. a binary star, as in the example with which this section is illustrated) is available. Such a model makes it easier to recognize those zeros (in the complex w-plane) that characterize the object.

6.4 Extending Zero-and-Add to Two Dimensions

The ZAA technique has been extended to permit processing of two-dimensional data in two somewhat different ways. Both methods involve reducing the two-dimensional data to a series of one-dimensional ensembles which are subsequently processed in the manner described in §6.2. The first extension, which is briefly discussed in §6.4.1, generates ensembles of one-dimensional integral projections (see §2.4.1) of the two-dimensional speckle images at several different angles. However, the second extension, which is mentioned in §6.4.2, employs zero-sheet concepts (Lane *et al.* 1987) that are described in §5.3.

Perhaps the most powerful implementation of ZAA would be realized by performing ZAA directly on the two-dimensional images. Such processing would involve generating the zero-sheets of the two-dimensional images and superimposing them, in the manner implied by (6.10). By appropriately compensating for the obdurate zeros, and thresholding the resultant zero-sheet, the quasi-steadfast zero-sheet could be identified and a reconstruction generated. This approach has an inherent advantage over the one-dimensional situation in that the zero-sheet of a two-dimensional object must be smooth and continuous. Because the zeros comprising a zero-sheet are necessarily connected it should be possible to identify sufficiently many zeros by tracking the zero-sheet. In contrast, the zero-maps of one-dimensional speckle images exhibit quasi-steadfast zeros at discrete locations on a plane i.e. they are unconnected. Consequently it is not then possible to invoke the connectedness of zero-sheets to assist when discriminating between obdurate zeros and quasi-steadfast zeros.

Unfortunately there are several, as yet insurmountable, technical problems associated with implementing this 'direct' two-dimensional ZAA. These problems largely arise because the zero-sheet of a two-dimensional image spans four-dimensional space. Thus, implementing processing on a digital computer in the manner described in

§6.2 seems, at present, prohibitively expensive, both in terms of storing the four-dimensional averaged zero-map, and in terms of the computational burden of generating the zero-sheet of each speckle image. In relation to tracking of a zero-sheet, it is informative to re-examine Fig. 5.10(b). These zero-contours of a typical image, of 32 × 32 pixel extent, provide an appreciation of the intricateness of a typical zero-sheet.

6.4.1 Using Projections

Sinton *et al.* (1986) indicate how to extend ZAA to two dimensions by forming one-dimensional projections (as defined by (2.35)) of the M two-dimensional speckle images. A series of projections, for projection angles ϕ_n; $n = 1, \ldots, N$, are made of each member of the original ensemble of two-dimensional speckle images. All the projections at each specific projection angle are collected separately to produces N ensembles of one-dimensional projections, each containing M members. The projection of the m^{th} speckle image at the n^{th} projection angle is written (*cf.* (2.35))

$$p_{s_m}^{\phi_n}(\xi) = p_f^{\phi_n}(\xi) \odot p_{h_m}^{\phi_n}(\xi) + p_{c_m}^{\phi_n}(\xi), \tag{6.19}$$

where each of the one-dimensional images is the integral projection at angle ϕ_n of its two-dimensional counterpart.

Consider now the ensemble of one-dimensional projections corresponding to $N = 1$. Since the object is identical in every two-dimensional speckle image, its projection is identical in every member of a particular ensemble of one-dimensional projections. However, the speckle psf changes with m, so its projection, in general, changes also. Thus, a particular ensemble of one-dimensional projections can be considered to be an ensemble of differently blurred projections of the object. Applying the one-dimensional ZAA technique described in §6.2 to this ensemble allows the thresholded, averaged zero-map $\bar{Z}_T\left\{p_{S_m(w)}^{\phi}\right\}$ to be generated. From this map $Z\left\{p_{F(w)}^{\phi}\right\}$ can be identified, and $p_{f(x)}^{\phi}$, the projection of the object at angle ϕ_1, can be reconstructed. By applying similar processing to each of the N ensembles of one-dimensional speckle images the one-dimensional projection of the object at each angle is generated.

Because reconstructing from zeros does not preserve the absolute amplitude of the image, the projections must be normalized. Normalization is conveniently achieved by weighting the pixels of each projection so that the pixels sum to unity.

The theory of compute(rise)d tomography (Garden 1984) reveals that an image can be reconstructed from an ensemble of integral projections. Several algorithms are available to achieve this, with perhaps the simplest being the backprojection algorithm (Herman 1979; Bates *et al.* 1983; Garden 1984). This technique involves backprojecting, or 'smearing', each of the one-dimensional projections across the two-dimensional plane at the angle at which the projection was obtained. The superposition of the N lots of backprojected data forms an estimate of the original two-dimensional object. This estimate is faithful provided a sufficiently large number of projections are computed, and that each projection is sufficiently faithful (Lewitt and Bates 1978; Lewitt *et al.* 1978). Thus, by backprojecting the reconstructions obtained from applying ZAA to the ensembles of one-dimensional projections, the two-dimensional image can be reconstructed.

6.4.2 Using Zero-sheets

Bates and Lane (1987b) (*cf.* Lane 1988, §6.4) generalize ZAA to two dimensions by application of the zero-sheet concept, introduced by Lane *et al.* (1987) (see §5.3).

In two-dimensions there is an infinite number of one-dimensional projections that can be reconstructed by fixing either ζ or γ in (5.15). As mentioned in §§6.2 and 6.4.1, it is often difficult, when the speckle images are contaminated, to accurately identify all of the zero clusters due to the object. Lane (1988, §6.4) notes that, in two dimensions, it is possible to form as many projections, and therefore as many zero-maps as are deemed necessary. The zero-maps are processed in the manner described in §6.2 and the locations of quasi-steadfast zeros are located. These zero positions are then processed in the manner of Curtis *et al.* (1985) to form a system of linear equations, which, when solved reconstructs the object. Curtis *et al.* (1985) note that the number of equations needed to solve for an $M \times M$ image is equal to $4M^2$.

This method has a distinct advantage over the method proposed in §6.4.1 in that every quasi-steadfast zero in a particular averaged zero-map need not be identified. The only critical constraint is that sufficient zeros be reconstructed in total. Thus, by choosing to compute a suitably large number of projections, usually only one or two zeros need be identified in each averaged zero-map. This means that the threshold applied to generate the thresholded zero-map can be increased, ensuring that fewer obdurate zeros are mistakenly identified as being truly steadfast. Thus, the susceptibility of the technique to contamination is reduced. (Bates and Lane 1987b) present a reasonably faithful reconstruction obtained by applying this technique to contaminated two-dimensional speckle images of a 5×5 object.

Chapter 7

Blind Deconvolution of Single Image

In many technical sciences a single image or signal, which is blurred in some manner, is recorded. Since one is usually interested in details of the unblurred object, it is highly desirable to be able to deblur the recorded data. When the exact nature of the blurring is known, one can readily invoke conventional deconvolution procedures, such as those discussed in §3.1, to implement the deblurring. However, in many situations the nature of the blurring is unknown, so the deblurring process in one of blind deconvolution (§1.5).

Algorithms for implementing the blind deconvolution of a single contaminated blurred image are the subject of this chapter. This general blind deconvolution problem is the most general and therefore the most difficult of all deconvolution problems. The conventional deconvolution algorithms reviewed in Chapter 3, the ensemble blind deconvolution algorithms discussed in Chapters 4 and 6, and the phase retrieval algorithms discussed in Chapter 5, can all be considered special cases of the general blind deconvolution problem.

Algorithms capable of implementing general blind deconvolution have only recently been reported, yet in this rapidly evolving area of information processing, several algorithms have already been realized. These algorithms are reviewed in this chapter, and a new technique is introduced and discussed in §7.4. The status of general blind deconvolution has recently been reviewed by Bates and Davey (1988).

Direct blind deconvolution (Lane and Bates 1987a), employing the concept of the zero-sheet (§5.3.1), is outlined in §7.1. Direct blind deconvolution is a natural extension of the direct phase retrieval technique of Lane *et al.* (1987), described in §5.3.3. This is unquestionably the most conceptually elegant and potentially powerful of all blind deconvolution techniques reported to date. Unfortunately, it is, as yet, only possible to usefully employ this method when a convolution is uncontaminated. Thus, this technique is currently of virtually no use for practical imaging processing, although the insight that zero-sheets provide into the question of uniqueness of more-than-one-dimensional blind deconvolution is extremely valuable.

In §7.2 an iterative technique for blind deconvolution of positive images is discussed and illustrated, and new applications for the technique are introduced. This technique comprises two separate stages, the first being the application of Fienup type phase retrieval algorithms (§5.2.2), and the second being the application of the modified magnitude retrieval algorithm (§5.4.2). This blind deconvolution technique was initially reported by Bates and Lane (1987a) (*cf.* Lane and Bates 1987b; Lane 1988) and has been applied in the context of astronomical speckle processing to remove the ghosting from SAA images (§4.8) (Bates and Davey 1988).

Ayers and Dainty (1988) have recently proposed a general blind deconvolution algorithm based upon an extended version of the general iterative loop presented in

Chapter 5. Their algorithm, which successfully deconvolves uncontaminated positive images, is outlined in §7.3.

A new blind deconvolution technique (Davey *et al.* 1988a), which can be appropriately considered as an adaptation of the technique of Ayers and Dainty (1988), is introduced and illustrated in §7.4. This algorithm is the first to have successfully deconvolved contaminated convolutions that are, in general, complex.

The application of general blind deconvolution techniques to deghosting shift-and-add images (§4.8) is described, and illustrated in §7.5.

A summary of the blind deconvolution techniques that are currently available is presented in §7.6. The efficacy of these algorithms, and the relative computational effort each requires to effect deconvolution, are compared in this section.

When describing general deconvolution algorithms in the following sections notation developed earlier in this thesis is employed. In particular, the contaminated isoplanatically blurred image $g(\mathbf{x})$ is written as

$$g(\mathbf{x}) = f(\mathbf{x}) \circledcirc h(\mathbf{x}) + c(\mathbf{x}), \tag{7.1}$$

where $f(\mathbf{x})$ is the true object, $h(\mathbf{x})$ is the psf describing the blurring, and $c(\mathbf{x})$, the *contamination*, incorporates any noise or other departures from the convolutional model upon which the imaging is based. Following the definition introduced in §1.5 the general deconvolution problem is posed as:

"Given only $g(\mathbf{x})$ recover $\hat{f}(\mathbf{x})$, an estimate of $f(\mathbf{x})$."

The material presented in §§7.1–7.3 is principally a review of the status of general blind deconvolution algorithms. The algorithm presented in §7.4 is new, and is especially significant since it is the first algorithm to successfully deconvolve a single contaminated convolution whose components are complex. The comparison of general blind deconvolution techniques is also original, as is the application of these techniques to the problem of deghosting shift-and-add images.

7.1 Direct Blind Deconvolution

It is now clear from the concept of the zero-sheet that any two-or-more-dimensional image, which is in general complex, and which is the convolution of N (a positive integer) individual images (called components), is in principle blindly deconvolvable. This means that, given merely the image, the value of N can be determined and each component can be recovered. Zero-sheets are discussed in Chapter 5, although the principle points which relate to blind deconvolution are discussed here.

The concept of the zero-sheet reveals that the zeros of the spectrum of each component of a convolution form a distinct K-dimensional surface in $2K$-dimensional space, where K denotes the dimensionality of the space spanned by the convolution. Consider the two-composite convolution $g(\mathbf{x})$, and its associated zero-sheet $Z\{G(\boldsymbol{\zeta})\}$. Since $g(\mathbf{x})$ is the convolution of $f(\mathbf{x})$ and $h(\mathbf{x})$, it follows that $Z\{G(\boldsymbol{\zeta})\} = Z\{F(\boldsymbol{\zeta})\} \cup Z\{H(\boldsymbol{\zeta})\}$ (*cf.* (5.20)). In §§5.3.3 and 6.1.1 it is noted that the task of deconvolving $g(\mathbf{x})$ is equivalent to partitioning $Z\{G(\boldsymbol{\zeta})\}$ into its two component zero-sheets $Z\{F(\boldsymbol{\zeta})\}$ and $Z\{H(\boldsymbol{\zeta})\}$. Once this partitioning is achieved, it is then possible to reconstruct the components from their zero-sheets (§5.3.2), thereby implementing blind deconvolution.

Lane and Bates (1987a) have shown that the algorithm for direct solution of the Fourier phase problem (Lane *et al.* 1987) (see §5.3.3), which is based on the concept of the zero-sheet, is directly extendable to solving the general blind deconvolution

problem in more-than-one dimension. By tracking along the zero-sheets, based on the property that they are analytic, and therefore must be smooth everywhere, it is possible to partition the zero-sheet of the convolution into the zero-sheets of the component images. Consider the situation for which $K = 2$, with the zero-sheets being two-dimensional surfaces embedded in four-dimensional space. As noted in §5.3.4, the intersections of such surfaces would be expected to intersect in zero-dimensional space, i.e. at points. However, the analyticity of the zero-sheets allows two intersecting zero-sheets to be unambiguously separated, as indicated in Fig. 5.13(a). (Lane and Bates 1987a) (*cf.* Lane 1988, §4.7) report deconvolving two 32x32 complex images from their uncontaminated 63 × 63 pixel convolution.

In the presence of contamination, the component zero-sheets lose their distinctiveness and 'bridge' at their points of intersection to form a single sheet, as discussed in §5.3.4 and illustrated in Fig. 5.13(b) and Fig. 5.15. Therefore, the zero-tracking algorithm, in its present form, is no longer able separate the component zero-sheets, and the deconvolution fails. This is why it is necessary to invoke iterative algorithms (§§7.2–7.4) to effect blind deconvolution in practice.

7.2 Modulo-π Phase Based Blind Deconvolution

The uniqueness of blind deconvolution, revealed by the concept of the zero-sheet, has stimulated the development of robust deconvolution techniques. These techniques are typically iterative, and tend to have been inspired by the iterative phase retrieval algorithms discussed in §5.2. The earliest of these techniques Bates and Lane (1987a) (*cf.* Lane and Bates 1987a; Lane 1988), which combines Fienup's phase retrieval algorithms (§5.2.2), with the modified magnitude retrieval algorithm (§5.4.2), is discussed in this section. This technique is henceforth called the modulo π phase based blind deconvolution technique.

When discussing the phase problem in Chapter 5 it was noted that only the image-form of an object can be recovered from its spectral magnitude. This limitation is normally of little account since the majority of relevant information about an image is incorporated in its image-form. However, consider the situation in which a composite image $g(\mathbf{x}) = f(\mathbf{x}) \odot h(\mathbf{x}) \longleftrightarrow F(\mathbf{u})H(\mathbf{u})$ is to be recovered from its known spectral magnitude $|G(\mathbf{u})|$. Unfortunately, from the point of view of phase retrieval, there is no unique image-form corresponding to $|G(\mathbf{u})|$. This follows since $\mathcal{P}[G(\mathbf{u})]$ is ambiguously related to $|G(\mathbf{u})|$ since

$$|G(\mathbf{u})| = |F(\mathbf{u})H(\mathbf{u})| = |F^*(\mathbf{u})H^*(\mathbf{u})| \tag{7.2}$$

and also

$$|G(\mathbf{u})| = |F(\mathbf{u})H^*(\mathbf{u})| = |F^*(\mathbf{u})H(\mathbf{u})|. \tag{7.3}$$

The inverse Fourier transforms of the products contained within the pairs of vertical lines shown in (7.2) have the same image-form, that of $f(\mathbf{x}) \odot h(\mathbf{x})$. However, the corresponding products in (7.3) have a different image-form, that of $f^*(-\mathbf{x}) \odot h(\mathbf{x})$. This ambiguity of the image-form inherent in the spectral magnitude of a composite image is illustrated in Fig. 7.1. Fig. 7.1(a) and (b) depict an object $f(x, y)$ and a psf $h(x, y)$, respectively, which are convolved to form $g(x, y)$ (Fig. 7.1(c)). When a phase retrieval algorithm is invoked to recover the image-form of $g(x, y)$ from its spectral magnitude, the algorithm can converge to any one of the four possible reconstructions shown in Fig. 7.1(c)–(f). Inspection of Fig. 7.1(c) and (d) reveals that these two reconstructions have the same image-form, that of $f(x, y) \odot h(x, y)$, whereas those shown in Fig. 7.1(e) and (f) have the image-form of $f^*(-x, -y) \odot h(x, y)$.

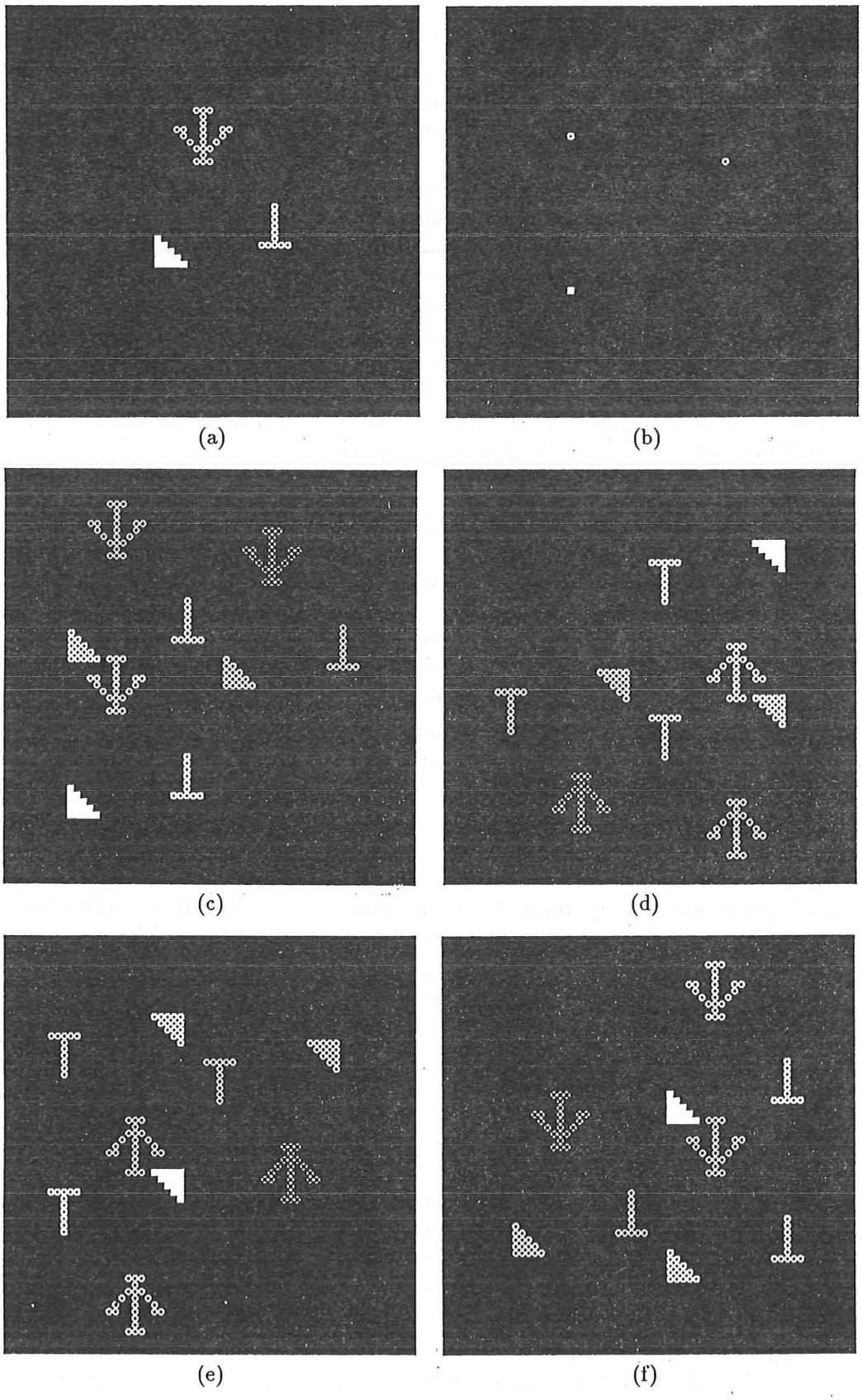


Figure 7.1: Ambiguous reconstructions to which a phase retrieval algorithm can converge when the image is composite: (a) $f(x,y)$; (b) $h(x,y)$; (c) $g(x,y) = f(x,y) \odot h(x,y) \longleftrightarrow F(u,v)H(u,v)$; (d) $f^*(-x,-y) \odot h^*(-x,-y) \longleftrightarrow F^*(u,v)H^*(u,v)$; (e) $f^*(-x,-y) \odot h(x,y) \longleftrightarrow F^*(u,v)H(u,v)$; (f) $f(x,y) \odot h^*(-x,-y) \longleftrightarrow F(u,v)H^*(u,v)$.

Bates and Lane (1987a) (*cf.* Lane 1988) have noted that the ambiguous reconstructions to which a phase retrieval algorithm can converge, when an image is composite, can be used to advantage to effect the blind deconvolution of the convolution of two real images. Since this blind deconvolution technique can only be applied when the components of the convolution are real, it is convenient to replace $f^*(-\mathbf{x})$ and $h^*(-\mathbf{x})$, with $f(-\mathbf{x})$ and $h(-\mathbf{x})$ respectively, in the following discussion.

The first step in deconvolving $g(\mathbf{x})$ by the technique of Bates and Lane (1987a) is to compute the convolution's spectrum $G(\mathbf{u}) = \mathcal{F}[g(\mathbf{x})]$. A Fienup phase retrieval algorithm (see §5.2) is then successively applied to the magnitude of the spectrum, i.e. $|G(\mathbf{u})|$, each time beginning with a different seed in the pseudo-random generator invoked to generate the random starting image. By starting the loop with several different random images, convergence to an image with the image-form $f(-\mathbf{x})\odot h(\mathbf{x})$ can be expected in at least one of the reconstructions. It is possible to ascertain when the reconstruction is of the image-form $f(-\mathbf{x})\odot h(\mathbf{x})$ because the spectral phase of such a reconstruction, $\Psi(\mathbf{u})$, differs significantly from $\Phi(\mathbf{u}) = \mathcal{P}[G(\mathbf{u})] = \mathcal{P}[F(\mathbf{u})H(\mathbf{u})]$. Thus, $\Psi(\mathbf{u})$ must correspond to either $\mathcal{P}[F(\mathbf{u})H^*(\mathbf{u})]$ or $\mathcal{P}[F^*(\mathbf{u})H(\mathbf{u})]$, although it is not possible to tell which it corresponds to.

The next step in the technique is the generation of the spectral phases of the $F(\mathbf{u})$ and $H(\mathbf{u})$. Addition and subtraction of $\Phi(\mathbf{u})$ and $\Psi(\mathbf{u})$ yields $2\mathcal{P}[F(\mathbf{u})]$ and $2\mathcal{P}[H(\mathbf{u})]$, although the ambiguity in $\Psi(\mathbf{u})$ means that it is not possible to know which of the resultant phases is that of $F(\mathbf{u})$ and which is that of $H(\mathbf{u})$. Unfortunately, these phases do not provide the true phase of $F(\mathbf{u})$ and $H(\mathbf{u})$ since conversion from, say, $2\mathcal{P}[F(\mathbf{u})]$ to $\mathcal{P}[F(\mathbf{u})]$ yields an ambiguity because

$$2\mathcal{P}[F(\mathbf{u})] \text{ modulo } 2\pi = 2\mathcal{P}[F(\mathbf{u}) \pm \pi] \text{ modulo } 2\pi. \tag{7.4}$$

Thus, it is only possible to determine the phase of $F(\mathbf{u})$ or $H(\mathbf{u})$ modulo π.

The final step in the modulo π phase based blind deconvolution technique is to invoke the modified magnitude retrieval algorithm described in §5.4.2. This algorithm reconstructs an image from its modulo π spectral phase. When invoking a magnitude retrieval algorithm to recover an image from its spectral phase, the support of the object must be known. Unlike Fourier phase retrieval, in which the support can be arbitrarily positioned, the support in magnitude retrieval must be correctly located since the spectral phase contains information about the position of the image. Lane (1988, §6.3) describes two schemes by which the magnitude retrieval algorithms can be invoked to recover the components of the convolution. The first scheme involves reconstructing both components simultaneously, applying the known $G(\mathbf{u})$ as an additional constraint ensuring correct convergence of the algorithm. If the support of either component is incorrectly estimated the algorithm can not converge. Thus, by re-applying the algorithm several times, each time with an estimate of the components' supports, the single occasion upon which the algorithm converges indicates that the supports are correct. However, (Lane 1988, §6.3) notes that, when one component is nearly symmetric, reconstructing both components simultaneously leads to poor convergence. He suggests a second, and preferrable, scheme that involves recovering one component from its modulo π phase for a number of different estimates of supports. When the extent of the support is too small the modified magnitude retrieval algorithm does not converge. However, when the extent of the support is too large the algorithm converges, but to a reconstruction comprising the true object convolved with a symmetric psf (see (5.24)). Only when the support is correctly estimated does the algorithm converge to the correct solution. It transpires, however, that it is possible to recognize that this 'first' reconstruction is not correct by Wiener filtering the known convolution with this reconstruction to generate a

'second' reconstruction, that of the other component. The presence of a symmetric psf in the first reconstruction must, essentially always, lead to severe degradation of the second reconstruction. Thus, only when the support is estimated correctly are both reconstructions faithful. Lane (1988, §6.3) notes that this latter scheme should be significantly more robust than the scheme in which both components are recovered simultaneously.

An example of deconvolution effected by the modulo π based blind deconvolution technique is displayed in Fig. 7.2. The object and psf are displayed in Fig. 7.2(a) and (b), and their convolution is depicted in Fig. 7.2(c). The spectral magnitude of this convolution was repeatedly applied to a Fienup phase retrieval algorithm until a reconstruction having the image-form of $f(x,y)\odot h(-x,-y)$ was reconstructed. This reconstruction is displayed in Fig. 7.2(d). The modulo π phases of the spectra of the two components were computed by adding and subtracting the spectra of Fig. 7.2(c) and (d). The modified magnitude retrieval algorithm was then applied to reconstruct the two components displayed in Fig. 7.2(e) and (f). These reconstructions compare very well with the original components of the convolution. A further example of modulo π phase based blind deconvolution is presented in §7.5.

The computational effort required to implement this modulo π phase based blind deconvolution technique is considerable. Firstly, Fienup processing must be repeatedly applied to $|G(\mathbf{u})|$ until a reconstruction having the image-form, $f(\mathbf{x})\odot h(-\mathbf{x})$, is obtained. Then an iterative magnitude retrieval algorithm must be invoked for all realistic combinations of object and psf support. When the extent of the object and psf are large, as many practical images tend to be, iterative loops containing multiple Fourier transformations tend to be computationally expensive. Furthermore, this technique requires several iterative loops to be performed at each stage, considerably compounding the required computational effort.

As a final note on modulo π based blind deconvolution, consider the special case in which it is known that the psf causing the blurring is symmetric. Since the spectral phase of such a psf can have only two values, zero or π (Hayes 1982), it follows that the modulo π phase of $H(\mathbf{u})$ is zero (Lane 1988, §2.5). Consequently the modulo π phase of the object can be obtained directly by computing

$$\begin{aligned} \frac{1}{2}\frac{G(\mathbf{u})}{G^*(\mathbf{u})} &= \frac{1}{2}\frac{|F(\mathbf{u})H(\mathbf{u})|e^{i\mathcal{P}[F(\mathbf{u})]+\mathcal{P}[H(\mathbf{u})]}}{|F(\mathbf{u})H(\mathbf{u})|e^{-i[\mathcal{P}[F(\mathbf{u})]+\mathcal{P}[F(\mathbf{u})]]}} \\ &= \mathcal{P}[F(\mathbf{u})] \text{ modulo } \pi. \end{aligned} \tag{7.5}$$

A modified magnitude algorithm can then be applied to $\mathcal{P}[F(\mathbf{u})]$ modulo π to reconstruct the object. Since symmetric psfs occur, at least to a good approximation, in many practical situations, such processing has significant potential applications including deblurring photographs that have been taken out of focus, or the camera has been 'bumped' whilst making an exposure (see Fig. 1.5). The application of such processing to practical photographic situations is currently being investigated by Hong Jiang at the University of Canterbury.

7.3 Ayers and Dainty's Algorithm

Ayers and Dainty (1988) have very recently proposed an algorithm for the blind deconvolution of a single blurred image. Their algorithm employs a single iterative processing loop to simultaneously recover both components of a convolution. By invoking a processing scheme based on this loop, the spectral phase of the convolution does not need to be discarded. Thus, their algorithm makes more complete use of the

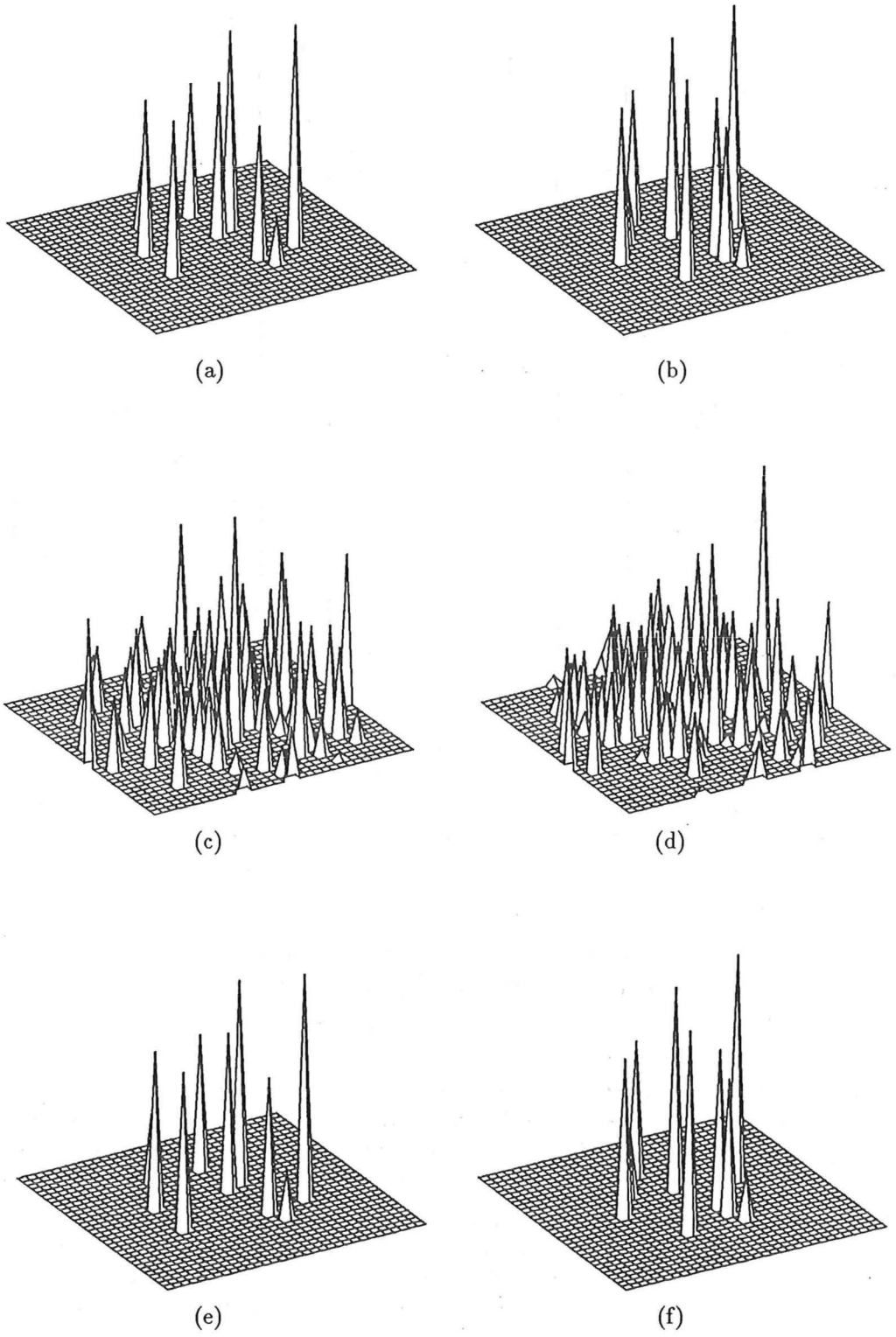


Figure 7.2: Blind deconvolution based upon modulo π phase: (a) $f(x,y)$; (b) $h(x,y)$; (c) $g(x,y) = f(x,y) \odot h(x,y)$; (d) reconstruction from phase retrieval having the image-form $f(x,y) \odot h(-x,-y)$; (e) reconstruction corresponding to $f(\mathbf{x})$; (f) reconstruction of $h(\mathbf{x})$.

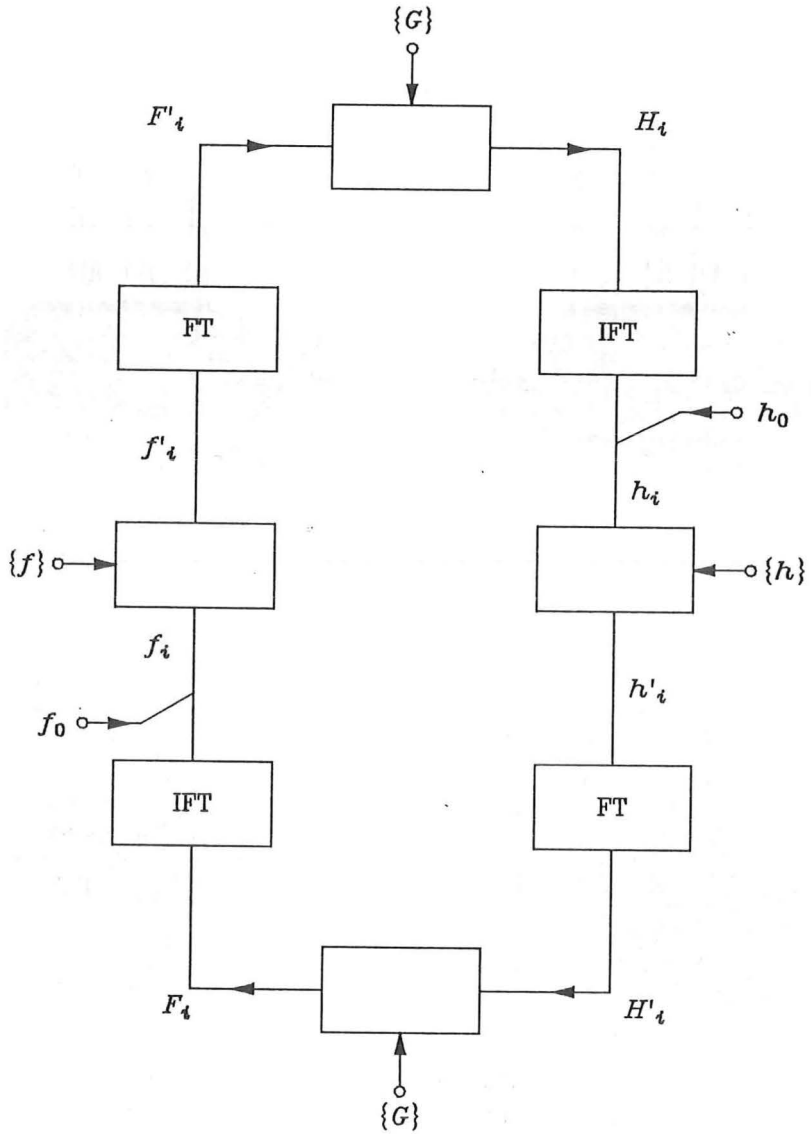


Figure 7.3: The extended iterative image retrieval loop proposed by Ayers and Dainty (1988). It is informative to compare this loop to the general iterative processing loop depicted in Fig. 5.2.

information contained in the convolution than the technique discussed in §7.2. Ayers and Dainty (1988) apply their algorithm to deconvolve positive images, although they note that the algorithm is conceivably extendable to bipolar and complex images.

The new processing loop introduced by Ayers and Dainty (1988), is here called the *extended iterative processing loop*, and is depicted in Fig. 7.3. This loop is conveniently considered as an extended form of the general iterative processing loop shown in Fig. 5.2, which is discussed in §§5.2 and 5.4 in the context of Fourier phase and magnitude retrieval algorithms. Each iteration of the extended iterative processing loop comprises two Fourier transformations, two inverse Fourier transformations, and the application of appropriate constraints interspersed with these transformations. The constraints attempt to 'urge' the algorithm to converge to the correct solution (Bates and Davey 1988).

The steps in the algorithm of Ayers and Dainty (1988) are now listed (refer also to Fig. 7.3):

(i) Compute $G(\mathbf{u}) = \mathcal{F}[g(\mathbf{x})]$.

(ii) Compute starting estimate $f_0(\mathbf{x})$ of $f(\mathbf{x})$ using a pseudo-random image, but also incorporating any *a priori* knowledge of $f(\mathbf{x})$.

(iii) Construct urged estimate $f_0'(\mathbf{x})$ of $f(\mathbf{x})$ by constraining $f_0(\mathbf{x})$ to satisfy the image-space constraint $\{f\}$.

(iv) Compute $F_i'(\mathbf{u}) = \mathcal{F}[f_i'(\mathbf{x})]$.

(v) Construct estimate $H_i(\mathbf{u})$ of $H(\mathbf{u})$, from $G(\mathbf{u})$, $F_i'(\mathbf{u})$ and $H_{i-1}(\mathbf{u})$.

(vi) Compute $h_i(\mathbf{x}) = \mathcal{F}^{-1}[H_i(\mathbf{u})]$.

(vii) Construct urged estimate $h_i'(\mathbf{x})$ of $h(\mathbf{x})$ by applying image-space constraints, $\{h\}$, to $h_i(\mathbf{x})$.

(viii) Compute $H_i'(\mathbf{u}) = \mathcal{F}[h_i'(\mathbf{x})]$.

(ix) Construct estimate $F_i(\mathbf{u})$ of $F(\mathbf{u})$, from $G(\mathbf{u})$, $H_i'(\mathbf{u})$ and $F_{i-1}(\mathbf{u})$.

(x) Compute $f_i(\mathbf{x}) = \mathcal{F}^{-1}[F_i(\mathbf{u})]$.

(xi) Construct urged estimate $f_i'(\mathbf{x})$ of $f(\mathbf{x})$ by applying image-space constraints, $\{f\}$, to $f_i(\mathbf{x})$.

(xii) If both $f_i'(\mathbf{x})$ and $h_i'(\mathbf{x})$ are sufficiently faithful according to some criterion, or if a preset number of iterations have been completed, proceed to step (xiv).

(xiii) Return to step (iv).

(xiv) Cease iterating and take $f_i'(\mathbf{x})$ and $h_i'(\mathbf{x})$ to be the best estimates of the object and psf, respectively.

It is informative to investigate the constraints that Ayers and Dainty (1988) apply in order to urge the algorithm to converge to the correct solution. The image-space constraint, applied in steps (vii) and (xi), is a positivity constraint (see §5.2). Any pixel of $f(\mathbf{x})$ or $h(\mathbf{x})$ that violates this constraint, i.e. it is negative, is set to zero. In addition to setting the negative pixels to zero, the image-space constraint proposed by Ayers and Dainty (1988) ensures that the sum of the pixel values of the estimate before and after the imposition of the constraint is the same. This is achieved by summing the negative values contained in the pixels that violate the positivity constraint, dividing this value by the number of pixels, and adding the resulting value to every pixel in the constrained estimate. Since this redistribution of the 'negative energy' may introduce further negative pixels into the constrained estimate, the image-space constraint is repeatedly reapplied until the constrained estimate is everywhere positive.

Consider now the implementation of the Fourier-space constraint outlined in steps (v) and (ix), in which an estimate of the spectrum of one of the components, say $F_i(\mathbf{u})$, is to be estimated from $G(\mathbf{u})$, $H_i(\mathbf{u})$ and $F_i(\mathbf{u})$. The most straightforward way to generate $F_i(\mathbf{u})$ is to invoke the simple inverse filter (§3.1.1) and divide $G(\mathbf{u})$ by $H_i(\mathbf{u})$. Ayers and Dainty (1988) point out that two difficulties exist with this processing (*cf.* §3.1.1). Firstly, the inverse filter produces erroneous results when the denominator possess pixels of very low magnitude (see, for example, Fig. 3.2(b)). Secondly, no information about the convolution is preserved at those spatial frequencies corresponding the zeros of $F(\mathbf{u})$ or $H(\mathbf{u})$. Both these limitations of the inverse

filter are discussed in §3.3. To counter these limitations, Ayers and Dainty (1988) propose the following rather intricate modification to the inverse filter.

When $|G(\mathbf{u})| <$ noise level

$$F_i(\mathbf{u}) = F'_{i-1}(\mathbf{u}) \tag{7.6}$$

else

$$\text{If } |H'_i(\mathbf{u})| \begin{cases} \geq |G(\mathbf{u})| & F_i(\mathbf{u}) = (1-\beta)F'_{i-1}(\mathbf{u}) + \beta\dfrac{G(\mathbf{u})}{H'_i(\mathbf{u})} \\ \text{otherwise} & \dfrac{1}{F_i(\mathbf{u})} = \dfrac{1-\beta}{F'_{i-1}(\mathbf{u})} + \beta\dfrac{H'_i(\mathbf{u})}{G(\mathbf{u})}, \end{cases} \tag{7.7}$$

where β is a positive constant in the range $0 \leq \beta \leq 1$, which is chosen before running the algorithm.

Ayers and Dainty (1988) note that the very low pixel magnitudes corresponding to most high frequencies in a typical convolution cause an inherent difficulty for deconvolution. Because their magnitudes are low, these frequencies are much more susceptible to contamination and it is considerably more difficult, if not impossible, to recover the information about the components corresponding to these frequencies. Ayers and Dainty (1988) counter this limitation by only reconstructing the spectrum within a restricted range of frequencies. To achieve this they introduce a Fourier-space apodization function, here denoted by $A(\mathbf{u})$, that is greater than zero for all frequencies up to some band-limit, and zero for frequencies beyond that limit. In practice, the actual band-limit is conveniently determined from $|G(\mathbf{u})|$. Additionally, $A(\mathbf{u})$ must be chosen such that $a(\mathbf{x}) = \mathcal{F}^{-1}[A(\mathbf{u})]$ is positive so that the positivity constraint is satisfied. In order to band-limit the estimate of a spectrum, the output of the modified inverse filter applied at steps (v) and (ix), is multiplied by $A(\mathbf{u})$. This ensures that pixels above the band-limit are set to zero. Additionally, Ayers and Dainty (1988) compensate for this spectral apodization by dividing $F'_i(\mathbf{u})$ and $H'_i(\mathbf{u})$, computed at steps (iv) and (viii) respectively, by $A(\mathbf{u})$.

Ayers and Dainty (1988) present two examples of reconstructions generated by their technique when operating upon uncontaminated convolutions of positive components. They report that, after selecting a suitable value for β, and applying an appropriate weighting function, the components of the convolution are recovered.

An example of a reconstruction effected by the algorithm of Ayers and Dainty is presented in Figs. 7.4 and 7.5. The object, $f(x,y)$, and psf, $h(x,y)$, for this simulation are depicted in Fig. 7.4(a) and (b), respectively. The convolution of these positive components is displayed in Fig. 7.4(c), while the starting estimate $f_0(x,y)$, is shown in Fig. 7.4(d). The reconstructions obtained after 500 iterations of the algorithm, with $\beta = 0.9$, are displayed in Fig. 7.4(e) and (f). Inspection of these reconstructions reveals that they correspond to the psf and object respectively. Since only a postivity constraint is applied, there is no necessity for the reconstructions to form at the centre of the image. Error-curves showing the violation of the image-space positivity constraint are displayed in Fig. 7.5(a). A similar reconstruction was attempted, but with contamination to the level of $\epsilon_{\text{dB}} = -20\text{dB}$ added to $g(x,y)$. Inspection of the resulting reconstructions, which are displayed in Fig. 7.4(g) and (h), reveals that the very small amount of contamination added to $g(x,y)$ for this simulation has affected the ability of this algorithm to recover the convolution's components. The resulting error-curves are displayed in Fig. 7.5(b).

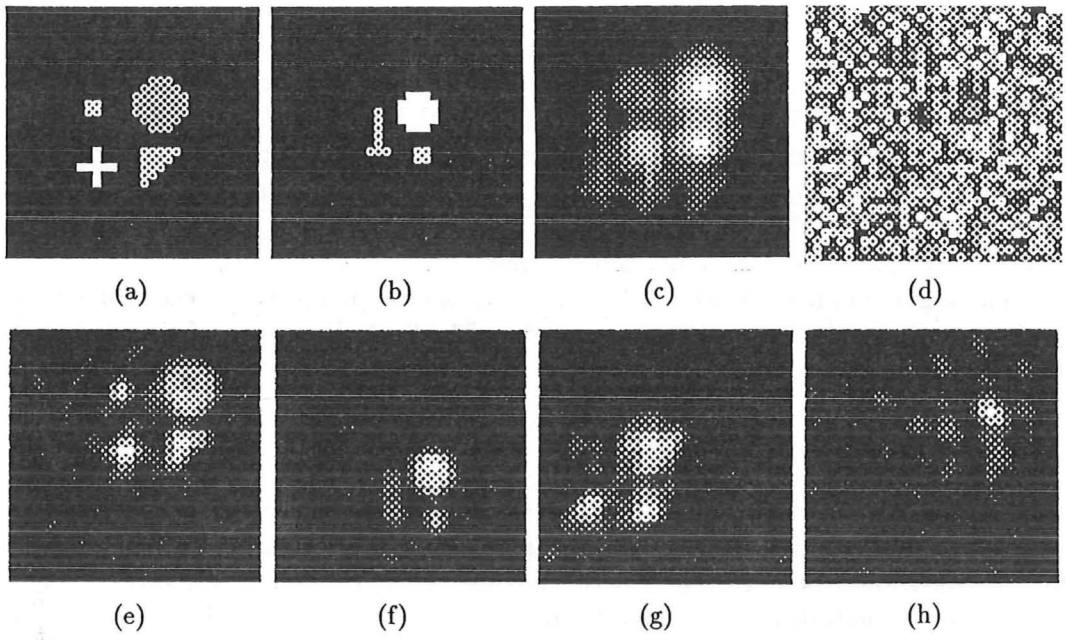


Figure 7.4: Deconvolution effected by the algorithm of Ayers and Dainty: (a) $f(x,y)$; (b) $h(x,y)$; (c) $g(x,y) = f(x,y) \odot h(x,y)$; (d) $f_0(x,y)$; (e) the first reconstruction corresponding to $h(x,y)$; (f) the second reconstruction corresponding to $h(x,y)$; (g) and (h) same as for (e) and (f), but with contamination to a level of $\epsilon_{\mathrm{dB}} = -20\mathrm{dB}$ added to the convolution.

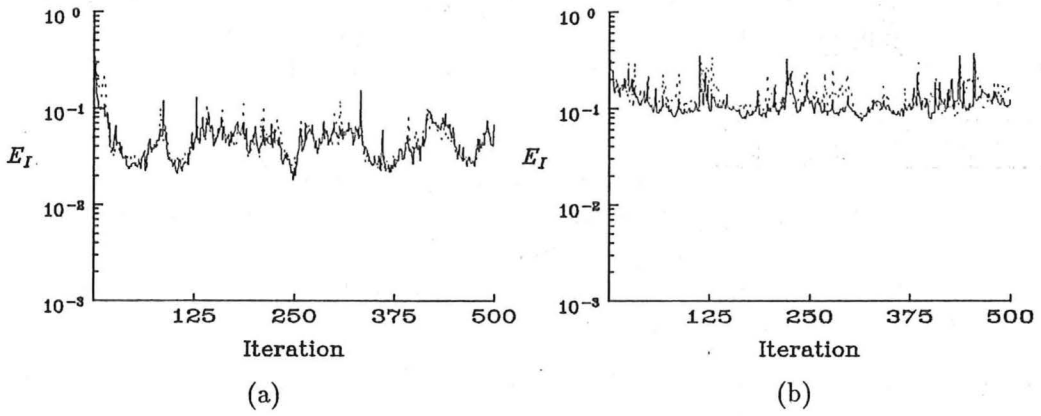


Figure 7.5: Error-curves for reconstructions shown in Fig. 7.4. The solid and dotted lines show the violation of the image-space constraint (*cf.* (5.11)) versus iteration number for the object and psf respectively: (a) no contamination; (b) $\epsilon_{\mathrm{dB}} = -20\mathrm{dB}$. Note the large fluctuations apparent in these curves.

7.4 A New Blind Deconvolution Algorithm

A new algorithm (*cf.* Davey *et al.* 1988a) for implementing the general blind deconvolution of a contaminated and blurred image is the subject of this section. This algorithm is best considered as an extension of the algorithm proposed by Ayers and Dainty (1988) (see §7.3), although it is interesting to note that both algorithms were under development concurrently. By applying different constraints than those applied by Ayers and Dainty, this algorithm can deconvolve a single contaminated image, that can, in general, be complex valued.

The algorithm is explained in §7.4.1. Results of applying it to deconvolve both positive and complex images are presented in §7.4.2. A discussion of the algorithm is presented in §7.4.3.

7.4.1 Algorithm Description

The new algorithm described in this section is based upon the extended iterative processing loop depicted in Fig. 7.3, but differs from that of Ayers and Dainty (1988) in the constraints that are applied to urge the iterative loop to converge. These constraints are now discussed with reference to Fig. 7.3 and the list of steps presented in §7.3.

7.4.1.1 Image-space constraints

The image-space constraints, which are applied at steps (xi) and (vii), are described here.

Provision is made for the optional imposition of each of the image-space constraints (i.e. positivity, reality and support) that are listed in §5.2. Any part of, say $f_i(\mathbf{x})$, that violates any of the applied constraints is set to zero in $f_i'(\mathbf{x})$. The actual constraints that are applied are determined from *a priori* knowledge of the imaging situation. For instance, if the convolution is formed by incoherent light, then the components are necessarily positive §2.1.2. Thus, the imposition of a positivity constraint is appropriate. Other situations may arise, for example in ultrasonic imaging (see §4.8.4), in which the components are complex. In such a situation only the support constraint is appropriate.

In order to apply a support constraint it is necessary to have an estimate of the support. Thus, when applying a support constraint in this blind deconvolution algorithm, in which two components are to be recovered simultaneously, it is necessary to have estimates of both components' supports. If these supports are not known *a priori*, estimates of them can be derived from the given convolution by the following procedure:

(i) The image-box, $B[g(\mathbf{x})]$ §2.1.4, of the convolution is obtained by appropriately thresholding $g(\mathbf{x})$.

(ii) A particular size for the image-box, $B[f(\mathbf{x})]$, which must be smaller than $B[g(\mathbf{x})]$ in each coordinate direction, is chosen. Any partial *a priori* information that may be available is incorporated into the choice of this image-box.

(iii) The extent of the image-box of the psf, $B[h(\mathbf{x})]$, which is compatible with $B[g(\mathbf{x})]$ and $B[f(\mathbf{x})]$, is then derived by invoking the extent of convolution theorem (2.27). As illustrated in Fig. 7.6, the combined length of the corre-

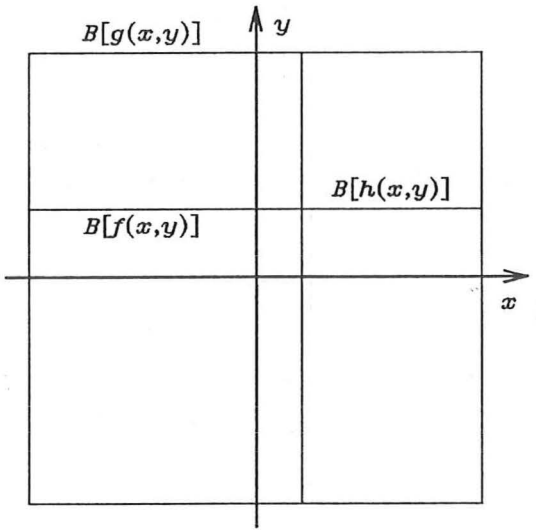


Figure 7.6: Illustration of the manner in which the extent of the image-boxes of the convolution's components determine the extent of the convolution's image-box. This illustrates the extent of convolution theorem (2.27).

sponding sides of $B[f(\mathbf{x})]$ and $B[h(\mathbf{x})]$ must equal the length of that side of $B[g(\mathbf{x})]$.

(iv) $B[f(\mathbf{x})]$ and $B[h(\mathbf{x})]$ are then located correctly relative to each other, and relative to $g(\mathbf{x})$ which is assumed to be centred at the origin of image-space. If $B[f(\mathbf{x})]$ is centred on the origin, then $B[h(\mathbf{x})]$ must also be centred. However, if $B[f(\mathbf{x})]$ is translated by $\mathbf{s}$ then $B[h(\mathbf{x})]$ must be translated by $-\mathbf{s}$ so that $f(\mathbf{x}) \odot h(\mathbf{x})$ is compatible with $g(\mathbf{x})$, as illustrated in Fig. 7.7. This is necessary since the image-space shifting property of the Fourier transform (Table 2.3) states that a translation of $\mathbf{s}$ in image-space produces a linear phase shift in the spectrum of $e^{i\mathbf{u}\cdot\mathbf{s}}$. Thus, if $f(\mathbf{x})$ is translated by $\mathbf{s}$, and $h(\mathbf{x})$ is translated by $-\mathbf{s}$, the phase shifts $e^{i\mathbf{u}\cdot\mathbf{s}}$ and $e^{i\mathbf{u}\cdot(-\mathbf{s})}$, in $F(\mathbf{u})$ and $H(\mathbf{u})$ respectively, cancel in $G(\mathbf{u})$. Consequently, $g(\mathbf{x}) = \mathcal{F}^{-1}[G(\mathbf{u})]$ is centred, as shown in Fig. 7.7(b). Conversely, if the components are not translated by opposite vector shifts, the resulting $g(\mathbf{x})$ is no longer centred, as depicted in Fig. 7.7(c). Applying this reasoning to the blind deconvolution algorithm, it follows that unless $B[f(\mathbf{x})]$ and $B[h(\mathbf{x})]$ are correctly positioned, the algorithm can not converge.

(v) The iterative blind deconvolution algorithm is applied using the image-boxes obtained from steps (ii) and (iv). The level of convergence of the algorithm is noted.

(vi) Steps (ii)–(v) are repeated for all sensible choices of the object image-box, $B[f(\mathbf{x})]$ (chosen at step (ii)). The combination of image-boxes, $B[f(\mathbf{x})]$ and $B[h(\mathbf{x})]$, which yield the most faithful reconstructions (according to the criterion introduced below) are taken to be the best estimates of the supports. The corresponding reconstructions provide the estimates of the components.

To evaluate the fidelity of the reconstruction one can compute the image-space error (5.11), which was introduced when discussing Fourier phase retrieval algorithms. However, the notation invoked here to describe this error differs slightly from that

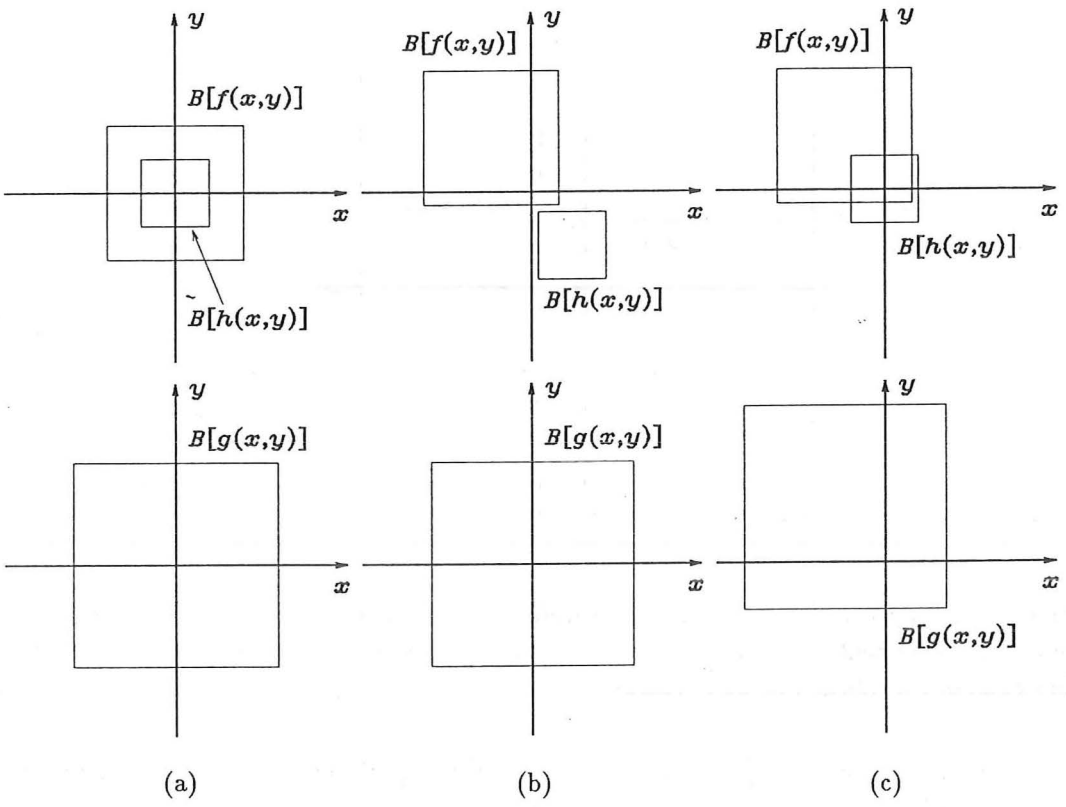


Figure 7.7: Illustration that, when one component of a convolution is displaced by s, the other component must be displaced by $-$s in order for the convolution's image-box to remain centred. This illustration shows how three different positionings of the components' image-boxes affect the position of $B\,[g(\mathbf{x})]$: (a) both $B\,[f(\mathbf{x})]$ and $B\,[h(\mathbf{x})]$ centred leads to $B\,[g(\mathbf{x})]$ centred; $B\,[f(\mathbf{x}-\mathbf{s})]$ and $B\,[h(\mathbf{x}+\mathbf{s})]$ results in $B\,[g(\mathbf{x})]$ centred; (c) $B\,[f(\mathbf{x}-\mathbf{s})]$ and $B\,[h(\mathbf{x})]$ leads to $B\,[g(\mathbf{x}-\mathbf{s})]$.

of Chapter 5, in that the image-space error at the i^{th} iteration for an image $q(\mathbf{x})$ is denoted by $E_{I,i}(q)$, where $q(\mathbf{x})$ represents either $f(\mathbf{x})$ or $h(\mathbf{x})$.

Alternatively, the true error (5.10) at the i^{th} iteration, which is here denoted by $E_{T,i}(q)$, can be computed. Of course the true error cannot be used in practice, because the true form of $q(\mathbf{x})$ is unknown *a priori*. However, the true error is useful for preliminary assessment of the algorithm's performance and this error measure is invoked to complement the image-space error for the computational examples presented in §7.4.2.

It is also appropriate to recall the concept of the contamination mask, introduced in §3.3, when evaluating the fidelity of the reconstructions obtained from a contaminated convolution. The known components are usefully band-limited by the contamination mask of the convolution, i.e.

$$q_{\Xi}(\mathbf{x}) = \mathcal{F}^{-1}\left[Q(\mathbf{u})\Xi\left[G(\mathbf{u})\right]\right], \tag{7.8}$$

where $Q(\mathbf{u})$ is either $F(\mathbf{u})$ or $H(\mathbf{u})$, and $\Xi\,[G(\mathbf{u})]$ is the contamination mask of the convolution for the particular level of additive contamination. $q_{\Xi}(\mathbf{x})$ forms a useful benchmark against which to compare $q_i(\mathbf{x})$. Henceforth in this chapter $E_{T,i}(q)$ is

computed as (*cf.* (5.10))

$$E_{T,i}(q) = \frac{\iint |q_{\Xi}(\mathbf{x}) - q_i(\mathbf{x})|^2 \, \mathrm{d}x \, \mathrm{d}y}{\iint |q_{\Xi}(\mathbf{x})|^2 \, \mathrm{d}x \, \mathrm{d}y}, \tag{7.9}$$

where $q_{\Xi}(\mathbf{x})$ is defined in (7.8).

7.4.1.2 Fourier-space constraints

The new algorithm's Fourier-space constraints, applied at steps (v) and (ix) in the list of steps presented in §7.3, are discussed here. These constraints are to be compared with the corresponding constraints applied in the algorithm of Ayers and Dainty (1988). Their algorithm invokes an intricate modification of the inverse filter, described by (7.6) and (7.7), in order to overcome the inherent limitations of a simple inverse filter (§7.3).

In comparison, the new algorithm described in this section invokes a modified form of the Wiener filter (§3.1.2) to implement these steps. In §3.3 it is noted that the Wiener filter can be viewed as a systematic method of improving poor reconstructions obtained when applying an inverse filter in the presence of contamination. Consequently, a Wiener filter allows a faithful reconstruction to be obtained, even in the presence of significant levels of contamination (see, for example, Fig. 3.4(a)).

The filter invoked in the new blind deconvolution algorithm is a modification of the traditional Wiener filter (3.8), and is described (for step (v)) by

$$H_i(\mathbf{u}) = \frac{G(\mathbf{u})F_i^*(\mathbf{u})}{|F_i(\mathbf{u})|^2 + \dfrac{\alpha}{|F_i(\mathbf{u})|^l}}, \tag{7.10}$$

where α is a positive real constant and l is an integer. Comparison with (3.8) reveals that this modification involves scaling the given Wiener filter constant by the magnitude of the quantity comprising the denominator at a particular pixel. The motivation for introducing such a modification is the desire to minimize the bias that is necessarily introduced by the Wiener filter. By scaling α by $|F_i(\mathbf{u})|$, the filter has most effect where this, the magnitude of the denominator, is small. Such scaling is appropriate since the inverse filter is most susceptible to contamination when the denominator is small (*cf.* §3.1.1). Conversely, when $|F_i(\mathbf{u})|$ is large, $\alpha/|F_i(\mathbf{u})|$ is small, so the bias introduced by the Wiener filter is reduced at these spatial frequencies. Note that when $l = 0$ and α is replaced by Φ, (7.10) reduces to (3.8).

Investigation of the convergence of the new blind deconvolution algorithm for several integer values of l, suggests that $l = 1$ produces acceptable convergence. Consequently, it is this value that is chosen for examples of the algorithm presented in this thesis. As usual, the Wiener filter constant, here specified in terms of the constant α, is conditioned by the contamination level, which effectively decreases as the iterations proceed. This follows because differences between $f_i(\mathbf{x})$ or $h_i(\mathbf{x})$ and $f(\mathbf{x})$ and $h(\mathbf{x})$, respectively, can be regarded as contributing to $c(\mathbf{x})$.

7.4.2 Results

Results of invoking the algorithm to deconvolve contaminated convolutions are presented here. The first example illustrates how critical the support constraint is, while the second confirms the ability of this algorithm to deconvolve contaminated complex blurred images.

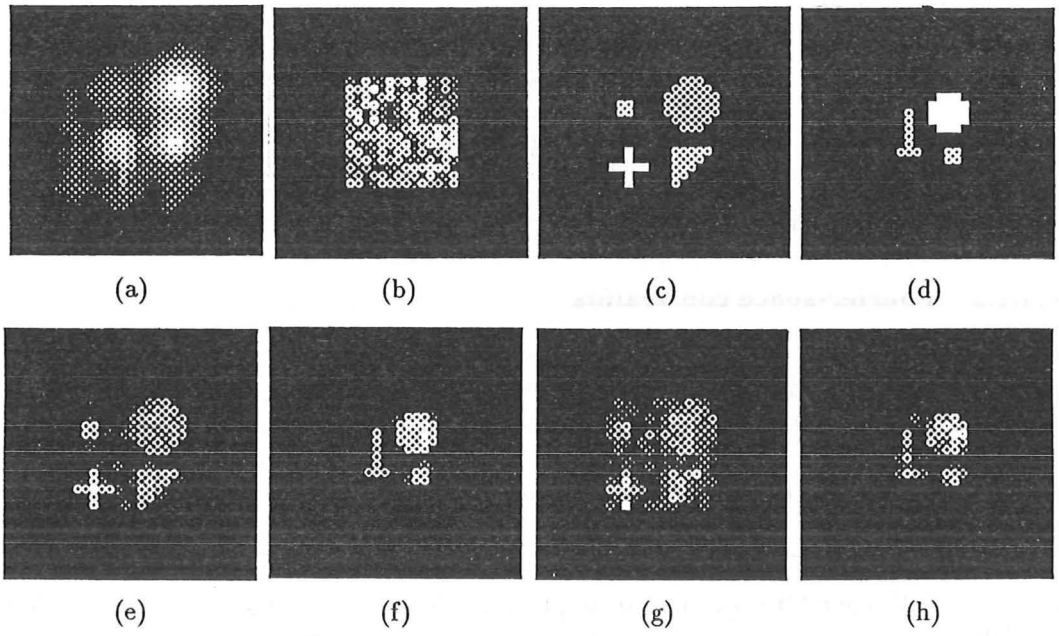


Figure 7.8: Blind deconvolution of positive blurred images: (a) $g(\mathbf{x})$ contaminated to a level of $\epsilon_{\text{dB}} = -10\text{dB}$; (b) $f_0(\mathbf{x})$; (c), (e), (g) $f(\mathbf{x})$: original, reconstructed for $\epsilon_{\text{dB}} = -30\text{dB}$, $\epsilon_{\text{dB}} = -10\text{dB}$; (d), (f), (h) $h(\mathbf{x})$: original, reconstructed for $\epsilon_{\text{dB}} = -30\text{dB}$ and $\epsilon_{\text{dB}} = -10\text{dB}$.

	ϵ_{dB}			
γ	$-\infty$dB	-30dB	-20dB	-10dB
1.0	0.032	0.034	0.048	0.27
1.2	0.034	0.039	0.061	0.54
1.4	0.40	0.41	0.51	0.55

Table 7.1: Values of the true error $E_{T,40}(f)$ obtained for pairs of values of γ and ϵ_{dB}, when applying the new blind deconvolution algorithm to the convolution shown in Fig. 7.8. $\alpha = 10^{-5}$ for the first 20 iterations and 10^{-6} for the final 20 iterations.

The blurred image deconvolved in the first example was generated by convolving the two positive images shown in Fig. 7.8(c) and (d), and adding contamination to one of four specified levels. The convolution corresponding to a contamination level of $\epsilon_{\text{dB}} = -10\text{dB}$ is depicted in Fig. 7.8(a). The blind deconvolution algorithm was then applied to each of these four contaminated convolutions, each time starting with the same random starting image, $f_0(\mathbf{x})$, shown in Fig. 7.8(b). Both positivity and support image-space constraints were applied. Additionally, the dependence of the algorithm's convergence on the correct support constraint was tested by applying the algorithm for three values of overestimation of the supports. To quantify this overestimation it is convenient to introduce a support overestimation factor γ that describes proportionally how much longer the sides of the image-boxes that are applied in the deconvolution are compared with the length of the sides of the actual image-box. Thus, for example, given that the true extent in the x coordinate direction of, say, $B[h(\mathbf{x})]$, is $L_h(x)$ (§2.1.4), the support that is applied in the blind deconvolution algorithm has an extent in that direction of $\gamma L_h(x)$. Such overestimation is likely

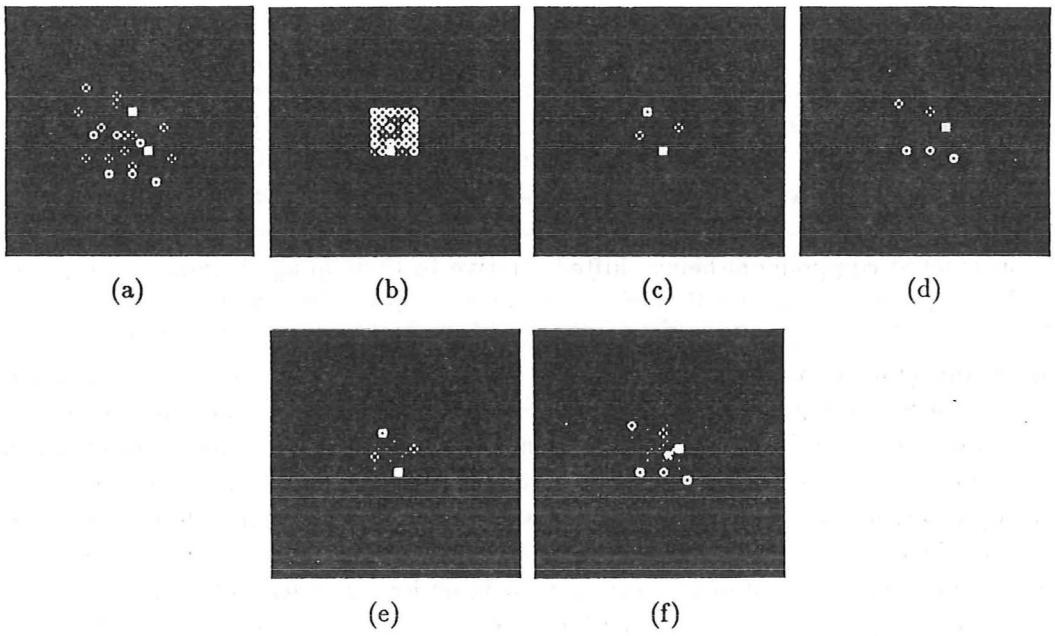


Figure 7.9: Blind deconvolution of a complex blurred image: (a) $|g(\mathbf{x})|$ for $\epsilon_{\mathrm{dB}} = -20\mathrm{dB}$; (b) $|f_0(\mathbf{x})|$; (c), (e) $|f(\mathbf{x})|$: original, reconstructed for $\epsilon_{\mathrm{dB}} = -20\mathrm{dB}$; (d), (f) $|h(\mathbf{x})|$: original, reconstructed for $\epsilon_{\mathrm{dB}} = -20\mathrm{dB}$.

to be unavoidable in practice, in the presence of appreciable contamination, since $B[g(\mathbf{x})]$ can be difficult to determine exactly.

Table 7.1 lists the values of $E_{T,40}(f)$ obtained for various pairs of values of γ and contamination level, when the contaminated convolutions of the components depicted in Fig. 7.8(c) and (d) are subjected to the new blind deconvolution algorithm. It is clearly advantageous to estimate the sizes of the boxes as accurately as possible. Fig. 7.8(e)–(h) show the reconstructions corresponding to $\gamma = 1$, for two contamination levels. Even when the contamination level is $\epsilon_{\mathrm{dB}} = -10\mathrm{dB}$, which is certainly appreciable, as inspection of Fig. 7.8(a) shows, the structure of the individual components is reproduced in the images reconstructed by the algorithm — see Fig. 7.8(g) and (h).

Fig. 7.9(c) and (d) depict the magnitudes of a pair of complex components, the phase of each of whose pixels was generated pseudo-randomly from a uniform distribution ranging from $-\pi$ to π. The magnitude of the convolution of these components, for a contamination level of -20dB, and the magnitude of the pseudo-random complex starting estimate, $f_0(\mathbf{x})$, are displayed in Fig. 7.9(a) and (b), respectively. Fig. 7.9(e) and (f) show the reconstructed components which are encouragingly faithful ($E_{T,150}(f) = 0.020$ and $E_{T,150}(h) = 0.026$). The algorithm successfully reconstructs the phases of the components. For these reconstructions the correct sizes of the image-boxes were assumed, and α was chosen to be 0.01.

7.4.3 Discussion

Blind deconvolution of a single blurred image, even one which is positive and uncontaminated, has only recently been implemented successfully. The new algorithm reported here is capable of recovering the positive components of a convolution when appreciable levels of contamination are added to the convolution. Additionally, this

is the first algorithm to successfully deconvolve a contaminated convolution having complex components. However, at present, it is only possible to successfully deconvolve rather simple complex images, of which those shown in Fig. 7.9 are typical.

Several practical difficulties exist with the implementation of this technique, however. For example, even when the components' supports are correctly estimated the algorithm may stagnate in some situations. Often this stagnation is due to the reconstructed components being shifted relative to their image-boxes, which necessarily results in some truncation of the reconstruction at the edges of the image-box. This cause of stagnation has also been identified in iterative Fourier phase retrieval algorithms (Fienup and Wackerman 1986; Lane 1988, §5.7), where this form of stagnation can be circumvented by re-positioning the image-box so that the component being reconstructed fits inside it. Experience has shown that, in situations in which truncation of the reconstructions occur, translating the components' image-boxes appropriately leads to superior reconstructions when implementing blind deconvolution. However, it is important to note that, if the support of, say, $f(\mathbf{x})$ is translated by $\mathbf{s}$, then the support of $h(\mathbf{x})$ must be translated by $-\mathbf{s}$ (refer to §7.4.1).

A further parameter that is crucial for the successful convergence of this algorithm is the Wiener filter constant, here specified in terms of α. For a given level of α, the tendency of this algorithm is to begin by converging rapidly. However, repeated iterations with the same value of α cause the rate of convergence to slow, and finally the algorithm starts to diverge. Experience has shown that before the algorithm starts to diverge, it is efficacious to reduce the value of α. This is sensible since the Wiener filter constant is conditioned by the contamination (§3.1.2). Furthermore, discrepancies between the true components and their current estimates are effectively lumped into $c(\mathbf{x})$. Thus, as the algorithm converges, the estimates become more faithful, so the effective magnitude of $c(\mathbf{x})$ decreases. It follows, therefore, that α should be reduced accordingly. Some method of dynamically varying α according to the fidelity of the reconstructions would be an appropriate avenue of ongoing research.

The uniqueness of multi-dimensional blind deconvolution, confirmed by the concept of the zero-sheet, provides an incentive for striving to make iterative general blind deconvolution algorithms more robust. It is appropriate to mention here the enormous increase in the effectiveness of iterative phase retrieval that eventuated from Fienup's introduction of the hybrid input-output algorithm (§5.2.2) to supplement his use of Gerchberg's error-reduction algorithm (§5.2.1) (which is perhaps only useful during the first few iterations). It seems sensible, therefore, to seek similar improvements to blind deconvolution, especially for complex images, so that they can be handled as efficiently as positive images exhibiting the same degree of pictorial detail.

7.5 Deghosting Shift-and-Add Images

The deghosting of shift-and-add images, effected by general blind deconvolution algorithms, is discussed in this section. This processing was foreshadowed in the discussion of shift-and-add imaging presented in §4.8.2.

Consider two ensembles of speckle images, recorded under (ideally) statistically identical seeing conditions, of resolvable and unresolvable objects, $f(\mathbf{x})$ and $\delta(\mathbf{x})$ respectively. The m^{th} images of each set are written as $s_m(\mathbf{x}) = f(\mathbf{x}) \circledcirc h_m(\mathbf{x}) + c_m(\mathbf{x})$ and $s_{\delta,m}(\mathbf{x}) = h_m(\mathbf{x}) + c_m(\mathbf{x})$ respectively.

Denoting $\mathbf{x}_m$ as that $\mathbf{x}$ for which $s_m(\mathbf{x})$ is greatest, the shift-and-add image of

$f(\mathbf{x})$ is (*cf.* (4.28))

$$f_{\mathrm{sa},M}(\mathbf{x}) = \langle s_m(\mathbf{x}+\mathbf{x}_m)\rangle_M, \tag{7.11}$$

where M is the number of speckle images of $f(\mathbf{x})$. Whatever shifting strategy is adopted, it necessarily transpires that

$$f_{\mathrm{sa},M}(\mathbf{x}) = f(\mathbf{x})\odot h_{\mathrm{sa}}(\mathbf{x}) + \tilde{c}(\mathbf{x}), \tag{7.12}$$

where $h_{\mathrm{sa}}(\mathbf{x})$ is the SAA psf (*cf.* (4.29)) and $\tilde{c}(\mathbf{x})$ incorporates all departures of $f_{\mathrm{sa}}(\mathbf{x})$ from the convolution model upon which the imaging is based. Similar processing applied to the speckle images of the unresolved star yields the image $f_{\delta,\mathrm{sa},M_\delta}(\mathbf{x})$, where M_δ is the number of speckle images of $\delta(\mathbf{x})$.

Although the iterative general blind deconvolution techniques described in this chapter can, in principle, recover an estimate of $f(\mathbf{x})$ directly from $f_{\mathrm{sa},M}(\mathbf{x})$, it is difficult to apply them efficiently because $f_{\mathrm{sa},M}(\mathbf{x})$ tends to fall off smoothly at its periphery, which inhibits (although it need not prevent) convergence of iterative image recovery. However, it is preferrable to begin by constructing a pre-processed image $\tilde{f}_{\mathrm{sa},M}(\mathbf{x})$ (*cf.* (4.35)), constructed by deconvolving $f_{\delta,\mathrm{sa},M_\delta}(\mathbf{x})$ from $f_{\mathrm{sa},M}(\mathbf{x})$ (*cf.* (4.35)), by invoking a Wiener filter (§3.1.2). $\tilde{f}_{\mathrm{sa},M}(\mathbf{x})$ is a faithful version of $f(\mathbf{x})$ only if the latter contains a dominatingly bright unresolvable point. In general, $\tilde{f}_{\mathrm{sa},M}(\mathbf{x})$ exhibits significant ghosting (§4.8.2), and can often be usefully expressed as (*cf.* (4.35))

$$\tilde{f}_{\mathrm{sa},M}(\mathbf{x}) = f(\mathbf{x})\odot h_{\mathrm{sa_o}}(\mathbf{x}) \tag{7.13}$$

where $h_{\mathrm{sa_o}}(\mathbf{x})$ is an object dependent 'ghosting' psf of unknown form, which can be estimated by invoking general blind deconvolution techniques.

This processing is now illustrated in Fig. 7.10 by the processing of speckle images of a multiple star (§4.3), which were computer generated in the manner described in §4.4.2. The actual multiple star, which comprises four individually resolvable components, is displayed in Fig. 7.10(a). A typical speckle image $s_m(x,y)$, resulting from the convolution of a randomly generated speckle psf and the object, Fig. 7.10(a), is displayed in Fig. 7.10(b). Although computer-generated, this speckle image has had contamination to a level of $\epsilon_{\mathrm{dB}} = -20\mathrm{dB}$ added. 256 such images were generated and processed by the basic SAA algorithm to form $f_{\mathrm{sa},256}(x,y)$, which is shown in Fig. 7.10(c). A further 256 speckle images of an unresolved star, $\delta(x,y)$, were generated and also processed by the basic SAA algorithm to provide $f_{\delta,\mathrm{sa},256}(x,y)$ (Fig. 7.10(d)). Wiener filtering $f_{\mathrm{sa},256}(x,y)$ with a Wiener filter derived from $f_{\delta,\mathrm{sa},256}(x,y)$ produced the 'defogged' SAA image, $\tilde{f}_{\mathrm{sa},256}(x,y)$, depicted in Fig. 7.10(e). Comparison of Fig. 7.10(a) and (e) clearly reveals that 'ghosts' are present in the latter image.

The iterative modulo π phase based blind deconvolution technique described in §7.2 was then invoked to deghost $\tilde{f}_{\mathrm{sa},256}(x,y) = f(x,y)\odot h_{\mathrm{sa_o}}(x,y)$, thereby providing an estimate of $f(x,y)$. Fienup phase retrieval (§5.2.2) was applied to $|\tilde{F}_{\mathrm{sa},256}(u,v)|$ repeatedly, each time commencing with a different random starting image, until a reconstruction having an image-form different from that of $\tilde{f}_{\mathrm{sa},256}(x,y)$ was obtained. The latter image-form necessarily corresponds to that of $f(-x,-y)\odot h_{\mathrm{sa_o}}(x,y)$, which is displayed in Fig. 7.10(f) (256 iterations of hybrid input-output were employed). Addition and subtraction of $\mathcal{P}\left[\tilde{F}_{\mathrm{sa},256}(u,v)\right]$ and the spectral phase of the reconstruction having the image-form of $f(-x,-y)\odot h_{\mathrm{sa_o}}(x,y)$ yielded the two modulo π phases, $\mathcal{P}[F(u,v)]$ and $\mathcal{P}[F_{sa_o}(u,v)]$, to which a modulo π magnitude retrieval algorithm (§5.4.2) was applied. In this simulation, adequate convergence was achieved when both components were reconstructed simultaneously (refer to §7.2). The algorithm was applied with several combinations of object and psf supports, with the

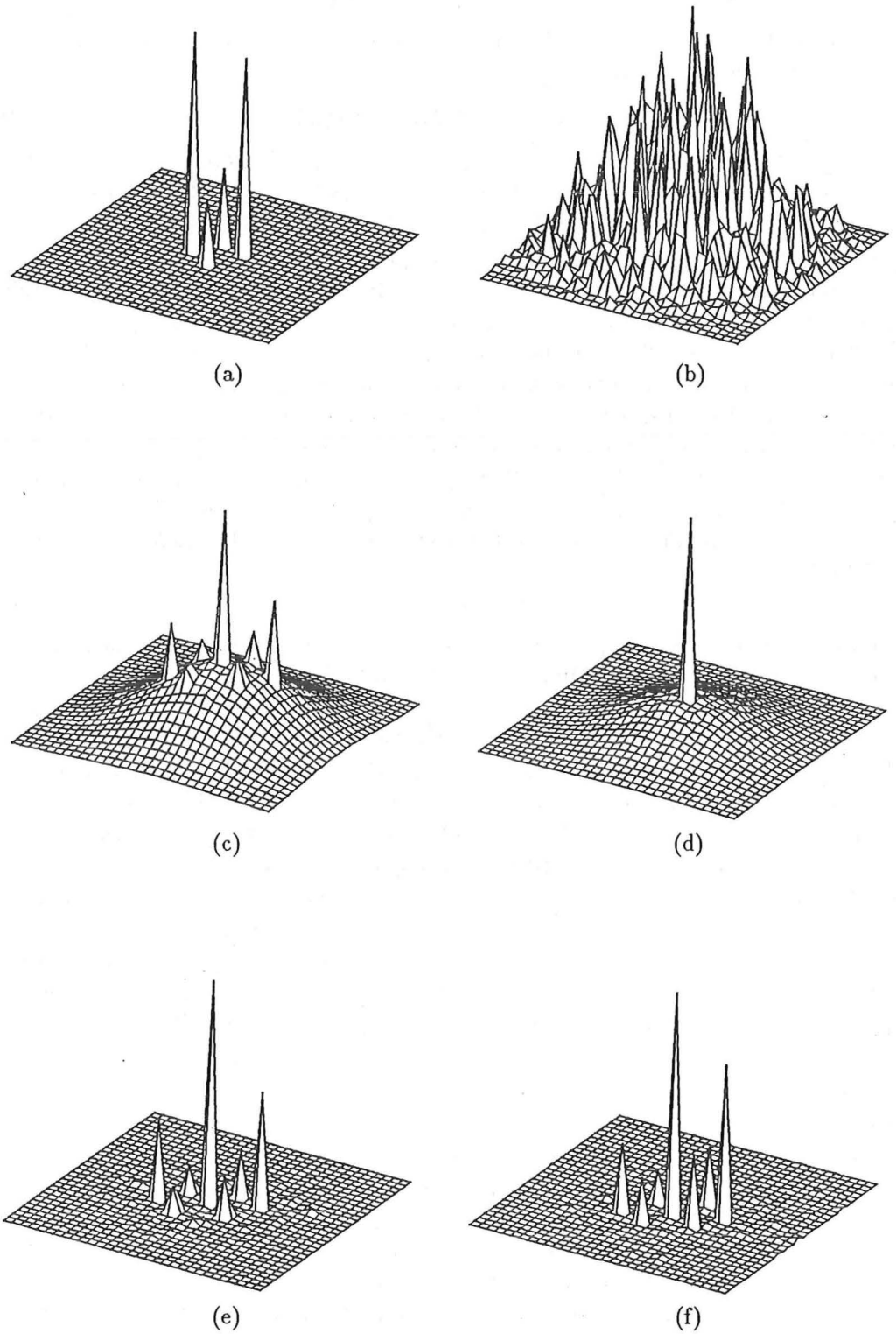


Figure 7.10: Deghosting the shift-and-add image of a multiple star by invoking the modulo π phase based blind deconvolution algorithm described in §7.2: (a) $f(x,y)$; (b) a typical speckle image $s_m(x,y)$; (c) the shift-and-add image $f_{\mathrm{sa},256}(x,y)$; (d) $f_{\delta,\mathrm{sa},256}(x,y)$; (e) $\tilde{f}_{\mathrm{sa},256}(x,y) \approx f(x,y) \odot h_{\mathrm{sa}_o}(x,y)$; (f) the convolution, having the image-form of $f(-x,-y) \odot h_{\mathrm{sa}_o}(x,y)$, resulting from applying Fienup phase retrieval algorithms to the spectral magnitude of (e). Note that Fig. 7.11 is a continuation of this figure.

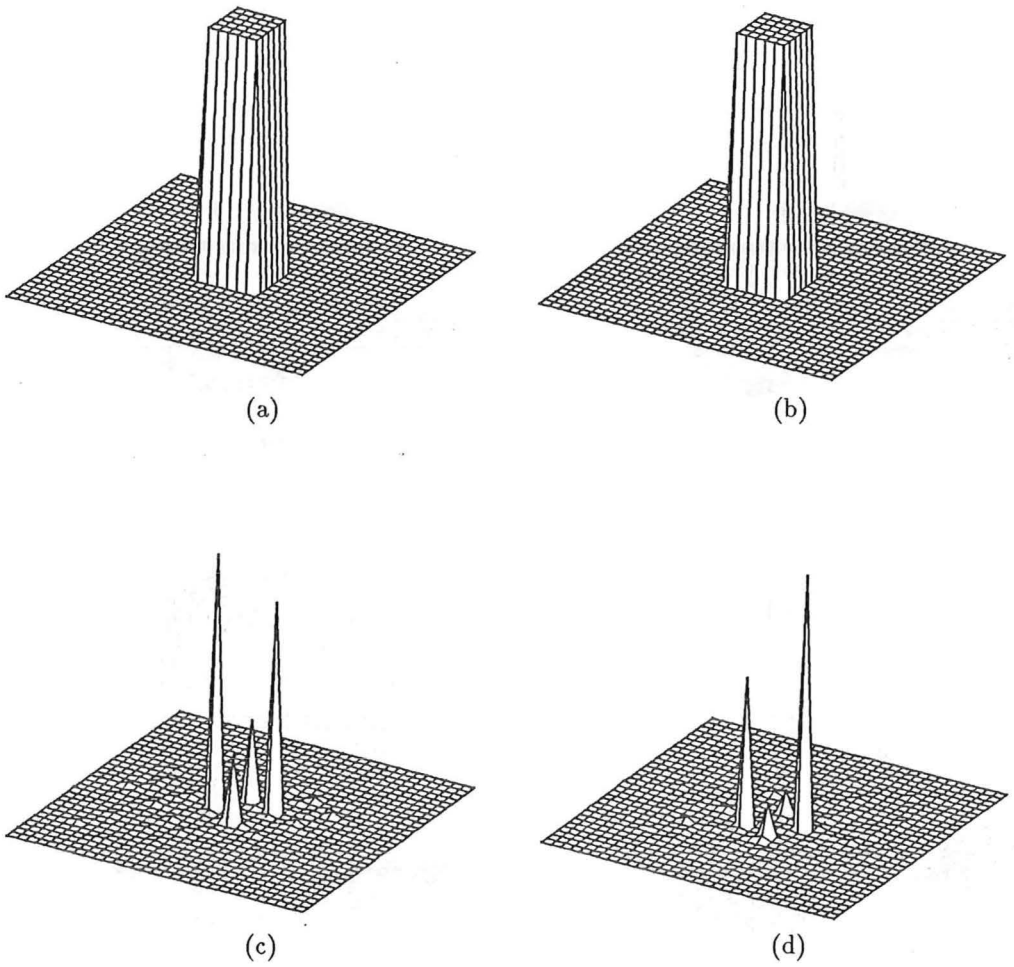


Figure 7.11: Deghosting a shift-and-add image by applying the blind deconvolution algorithm which is based on the modulo π phase (continuation of Fig. 7.10): (a) and (b) the supports for which the algorithm converged; (c) and (d) the reconstructions. Recalling the nature of the formation of SAA ghosts makes it apparent that (c) and (d) correspond to $f(x,y)$ and $h_{\mathrm{sa}_o}(x,y)$, respectively.

combination of supports that produced convergence displayed in Fig. 7.11(a) and (b). The reconstructions resulting from the magnitude retrieval algorithm (when these supports were applied) are presented in Fig. 7.11(c) and (d). Upon comparing the object Fig. 7.10(a) with these two reconstructions, it is apparent that Fig. 7.11(c) corresponds to $f(x,y)$, and is clearly very faithful, given the level of contamination applied to the individual speckle images. Fig. 7.11(d) corresponds to $h_{\mathrm{sa}_o}(x,y)$, which basically has four non-zero pixels, located at mirror inverse positions of the non-zero pixels in the object. In practice, it may on occasion be somewhat more difficult to determine which of the two reconstructions corresponds to the object, since $f(x,y)$ is not available. However, the ambiguity can always be readily resolved from knowledge of the mechanism of ghost formation.

Several refinements of this deghosting technique have been devised to make the technique more robust and also more computationally efficient. The first refinement involves determining an estimate of the support of the object in order to reduce the

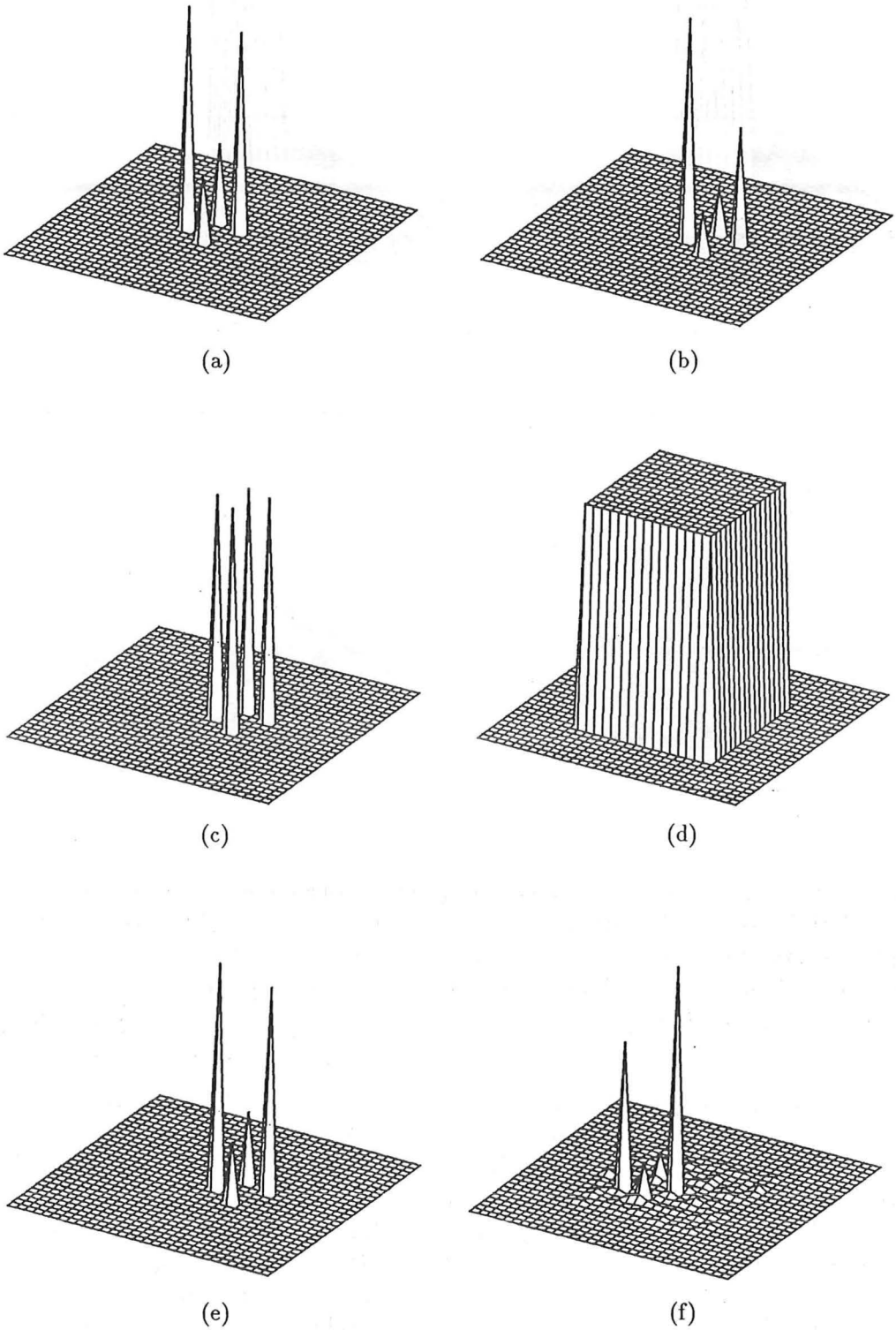


Figure 7.12: Deghosting the shift-and-add image displayed in Fig. 7.10(e) by invoking the general blind deconvolution technique described in §7.4: (a) $f(x, y)$ (the same as Fig. 7.10(a)); (b) $f_0(x, y)$ derived from (a); (c) the exact support of the object; (d) an estimate of the support of $h_{\mathrm{sa}_o}(x, y)$; (e) the reconstruction obtained from the support shown in (c), corresponding to $f(x, y)$; (f) the reconstruction from the support shown in (d) ($h_{\mathrm{sa}_o}(x, y)$).

number of times a blind deconvolution technique must be applied to test for convergence. The image-box of the object can often be obtained accurately, either directly from the SAA image, or by some ancillary processing (*cf.* Sinton *et al.* 1986; Bates and Davey 1987b). For example, speckle interferometry (§4.5) can be applied to the speckle images to generate the autocorrelation of the object, $ff(x,y)$. By appropriately thresholding $ff(x,y)$, the image-box $B\,[ff(x,y)]$ can be determined. It follows from the extent of autocorrelation theorem (2.28) that the extent of $B\,[ff(x,y)]$ in each coordinate direction is twice the extent of $B\,[f(x,y)]$. Thus, $B\,[f(x,y)]$ can readily be deduced as a rectangle whose sides have half the extent of those of $B\,[ff(x,y)]$. Knowledge of the image-box reduces the computational effort required to implement the deghosting, since the number of combinations of support that must be applied to the magnitude retrieval algorithm is drastically reduced. In fact, when employing the techniques introduced in §§7.3 and 7.4, only a single combination of the components' supports need be tried. When employing the technique introduced in §7.2, however, the supports must be tried in different locations.

Furthermore, it is often possible to determine a more exact estimate of the support of certain classes of object than the simple rectangular image-box. Consider for example, the object displayed in Fig. 7.10(a), which comprises four separated unresolvable components. For such an object the relative positions of the components can often be accurately obtained from inspection of either $f_{\mathrm{sa},M}(x,y)$ or the speckle interferometric image, even though the relative magnitudes may not be accurately deducible. Thus, for the blind deconvolution of the $\tilde{f}_{\mathrm{sa},M}(x,y)$ shown in Fig. 7.10(e), it is possible to conclude that the object support consists of four pixels having the relative juxtaposition of the peaks apparent in Fig. 7.10(a). An accurate estimate of the support accelerates convergence, and also improves the robustness, of the technique.

An additional refinement, which accelerates convergence when applying iterative blind deconvolution algorithms, is to start the iterations with an estimate of $f(x,y)$ rather than a random image. Such an estimate is conveniently obtained from $\tilde{f}_{\mathrm{sa},M}(x,y)$ by moving the image-box, $B\,[f(x,y)]$, over the defogged SAA image, $\tilde{f}_{\mathrm{sa},M}(x,y)$, until the sum of all pixel values within the box is a maximum. The contents of the box at this location then represent the best available estimate of the true object, which can be usefully taken as $f_0(x,y)$.

An example of deghosting the SAA image shown Fig. 7.10(e), by applying the blind deconvolution algorithm described in §7.4, is displayed in Fig. 7.12. For these reconstructions, iterations commenced with the estimate of the object (displayed in Fig. 7.12(b)) that was extracted from $\tilde{f}_{\mathrm{sa},M}(x,y)$ in the manner described above. The estimates of the supports of $f(x,y)$ and the ghosting psf were also deduced from $\tilde{f}_{\mathrm{sa},M}(x,y)$ and are displayed in Fig. 7.12(c) and (d) respectively. After only 30 iterations of the blind deconvolution algorithm described in §7.4, the reconstructions of $f(x,y)$ and $h_{\mathrm{sa}_0}(x,y)$, shown in Fig. 7.12(e) and (f) respectively, were obtained. The very small number of iterations required to reconstruct the object highlights the expediency of incorporating the abovementioned refinements into the SAA deghosting procedure.

Chapter 8

Conclusions and Recommendations for Future Studies

This thesis reviews existing algorithms, and presents new algorithms, for processing degraded information arising in two quite separate fields of scientific endeavour. In this chapter conclusions from the work presented in this thesis are drawn, aspects of this work warranting further investigation are identified, and avenues for future research are suggested.

8.1 Deconvolution

8.1.1 Conclusions

Convolution of two images is equivalent to multiplication of their spectra, so deconvolution is equivalent to factorization of their spectra. The analytically continued spectrum of a K-dimensional image is $2K$-dimensional and is zero on a $(2K-2)$-dimensional zero-sheet. When $K=1$ it is always possible to factorize a polynomial, with the factors corresponding to isolated points on the complex plane where the z-spectrum is zero. Consequently there is no unique solution when deconvolving a one-dimensional image. However, when $K \geq 2$, a z-spectrum is only factorizable, or reducible, if it is a convolution, with the factors corresponding to the components of the convolution.

Deconvolution techniques can be split into four categories. These are conventional blind deconvolution techniques in which an estimate of the blurring of a single blurred image is available, ensemble blind deconvolution in which an ensemble of differently blurred versions of the same image is available, and general blind deconvolution techniques when a single blurred image is available.

Conventional blind deconvolution algorithms are available for situations in which one knows the exact nature of the blurring. The inverse filter provides an adequate reconstruction when the psf is known exactly, and the imaging situation is contamination free. Algorithms such as the Wiener filter and CLEAN can reconstruct palpably faithful images in the presence of considerable contamination.

Phase retrieval is a particular subclass of blind deconvolution, in which the psf is the conjugate mirror image of the object. Iterative algorithms implementing phase retrieval have proven robust in the presence of considerable contamination.

The concept of the zero-sheet, introduced by Lane *et al.* (1987), provides an elegant proof of the uniqueness of phase retrieval in more-than-one dimensions, which can be readily extended to encompass general blind deconvolution. Direct phase retrieval and blind deconvolution schemes (Lane and Bates 1987a), based upon sepa-

rating the zero-sheets, are effective in the absence of contamination. However, when a convolution is contaminated the component zero-sheets bridge and the direct deconvolution fails. However, the zero-sheet concept motivates the pursuit of robust iterative blind deconvolution algorithms, capable of deconvolving a single contaminated blurred image.

Zero-and-add (Chapter 6) represents an approach to ensemble blind deconvolution that is fundamentally different from all other ensemble blind deconvolution techniques (see Chapter 4). When invoked upon effectively isoplanatic one-dimensional astronomical speckle images, ZAA is capable of faithfully imaging arbitrary objects without the need to correct for the effects of astmospheric seeing. However, various forms of contamination influence the quasi-steadfast zeros in the ZAA map in an, as yet, little understood manner. In particular, windowing of the speckle images seems to produce severe distortion in the locations of the quasi-steadfast zeros. ZAA can be extended to process two-dimensional speckle images by reducing these images to a series of ensembles of one-dimensional images. The one-dimensional ZAA technique is then separately applied to each ensemble, with a two-dimensional reconstruction of the object obtained from the series of one-dimensional reconstructions.

A composite speckle imaging technique (§6.3), for processing one-dimensional infrared speckle images, can recover the image with significantly less computational effort than alternative speckle imaging algorithms. This technique combines several image-processing algorithms, including SAA and ZAA, to extract the object from infrared speckle images. It also automatically compensates both for limitations of the recording equipment and for atmospheric seeing effects.

Blind deconvolution techniques have recently been realized for deconvolving a single blurred image. A new iterative algorithm, introduced in this thesis, is the first to have effected the deconvolution of a single contaminated convolution of complex components. As yet, only the deconvolution of relatively simple complex components has been achieved.

8.1.2 Future Research

In order to apply astronomical speckle imaging techniques with confidence, one must recognize the strengths and weaknesses of the various algorithms when the speckle images are non-ideal. Thus, detailed comparisons of all speckle imaging techniques, similar to those of, for example, Ayers *et al.* (1988) or Freeman *et al.* (1987), are required. Bates and Davey (1987a) outline several criteria upon which a comparison of speckle imaging techniques could be based.

One such comparison of astronomical speckle imaging algorithms, which would have particular significance to the work presented in this thesis, would be to compare the performance of the Fork algorithm (§4.8.3), with that of the shift-and-add (SAA) algorithm coupled with the deghosting procedure described in §7.5. The Fork algorithm is predicated upon knowing, firstly, that a binary is being observed, and secondly, the separation and position angle of that binary. Since this information is equivalent to knowing the exact support of the object, it would be appropriate to apply the exact support in the SAA deghosting procedure.

Perhaps the most promising avenue of future research arising from zero-and-add is true two-dimensional ZAA as foreshadowed in §6.4. Applying ZAA to two-dimensional images, as it is presented in §§6.4.1 and 6.4.2, involves reducing the ensemble of two-dimensional images to a series of one-dimensional images to which ZAA is applied. The zeros common to the spectra of the one-dimensional images are then identified and a series of reconstructions obtained, from which the two-dimensional

reconstruction can be generated. In the presence of considerable contamination, these quasi-steadfast zeros are often difficult to locate, and a degraded reconstruction results. An inherently more robust technique would be realized by identifying sections of the zero-sheets that are common to the zero-sheets of each of the images. Since the zeros comprising a zero-sheet form a continuous surface, the common sections of the zero-sheets should be able to be located more accurately than the zeros of the spectra of one-dimensional images. Implementing this two-dimensional processing introduces technical difficulties, however, since the zero-sheets are two-dimensional surfaces embedded in four-dimensional space. Consequently, storing and effectively processing such images is, as yet, beyond the current capability of most computer systems. However, advances in progamming techniques and computer hardware may overcome these difficulties.

Many opportunities for future research arise from the iterative blind deconvolution algorithm presented in §7.4. Many of these possiblities relate to schemes for improving the convergence properties of the algorithm. One such area requiring further investigation is a mechanism for automatically choosing α according to the fidelity of the reconstruction at a particular iteration. Further investigation is also required into the optimal form of the Wiener filter for this algorithm. Preliminary evaluation has indicated that the modified Wiener filter, described by (7.10), provides superior algorithm convergence than the more traditional Wiener filter (3.8). Furthermore, additional research is needed to investigate the harmful effect of the bias introduced by any Wiener filter (see §3.3). It is interesting to note that the blind deconvolution algorithm of Ayers and Dainty (1988) avoids the bias introduced by the Wiener filter, but it appears to be somewhat less robust in the presence of contamination.

8.2 Neurology

8.2.1 Conclusions

A computer based system for automated detection of epileptiform activity in the EEG has been developed. The system consists of two distinct stages. The first is a feature extractor that utilises parts of previously published spike-detection algorithms to produce a list of all spike-like occurrences in the EEG. The second, expert system, stage employs rules incorporating knowledge elicited from an electroencephalographer (EEGer) to discern which of the possible spikes and sharp-waves (SSWs) detected by the feature extractor represent true epileptiform activity. The output from the algorithm is a summary of the detected epileptiform events, which is available to the EEGer in interpreting the EEG.

The performance of the expert system is compared with an EEGer using a 320 second segment from an EEG containing epileptiform activity. The system detected 19 events, and missed 7 (false negative), which the EEGer considered epileptiform. There were no false positive detections. These provisional results indicate the system's potential for reliably detecting epileptiform activity.

The results of the algorithm are quoted in terms of the number of epileptiform events that the system identified, rather the number of individual SSWs detected. This method of evaluating the system's performance is more meaningful, since an EEGer is usually more interested in the correct detection of epileptiform events, rather than the detection of individual SSWs. The EEGer is, however, interested in the channel upon which the spikes comprising the epileptiform event occur, since that information often allows the site at which the activity arises to be identified. Thus, the individual channels which provide evidence for each epileptiform event must be

available to the EEGer.

Although many attempts have been made in the last two decades to automate the detection of epileptiform activity (*cf.* Ktonas 1983; Gotman 1985), only moderate detection accuracy has been achieved. The system reported in Appendix B is the first to apply expert system techniques to this task. Expert systems are inherently suited for automatic detection of epileptiform activity in EEG since experts, who routinely read EEGs, are available in all large medical centres. Furthermore, the knowledge about epileptiform activity, which is elicited from such an expert, can often be readily translated into rule form.

The task of accurately eliciting knowledge from the expert is fundamentally important to the success of an expert system. It is, however, often not an easy task since many experts have difficulty expressing their expertise, which has often been acquired over many years, from a wide variety of sources.

8.2.2 Future Research

The system described in Appendix B is a prototype system, and, as such, has not been developed to a sufficiently sophisticated and 'user-friendly' level for routine application in a clinical environment. Re-implementation of the system, in a format designed for routine use by clinicians, who are not expert computer operators, is under development by Dingle (1988) in conjunction with colleagues at Christchurch Hospital. This new implementation is based on an IBM personal computer, with software written in C and PROLOG.

One limitation of the system, as described in Appendix B, is that only rules for the montage (i.e. electrode configuration) which is most commonly used in the Neurology Department at Christchurch Hospital have been implemented. However, when developing the system, explicit provision has been made for adding further rules that incorporate the expected relationships between SSWs occurring on various channels when recording the EEG with other montages.

The production system approach has a distinct advantage over other, more conventional programming techniques, in that extra rules can be readily added to the system. Thus, additional information elicited from EEGers in the future can be rapidly encapsulated into the system.

An interesting research possibility would be to combine the spike detection system with the shift-and-add algorithm in order to investigate trends in the EEG that may accompany epileptiform activity. This processing would involve invoking the spike detection system to identify SSWs, which would then be taken as the shift-and-add reference points for the subsequent application of the SAA algorithm. Each channel of EEG would be shifted so that the epileptiform event lies at the origin. In addition, the polarity of the SSW (i.e. does it point up or down) would have to be tested, because it is likely that some of the EEG channels would need to be inverted. Several strategies for implementing the 'adding' part of shift-and-add are possible. For instance, one strategy would involve separately summing all the SSWs that occur, at different times, on each channel. Provided that sufficiently many SSWs are averaged, one would expect the ongoing EEG to cancel, with the remaining averaged waveform representing the epileptiform activity. Such processing would reveal any characteristic features marking the onset or termination of epileptiform activity on a particular channel. Alternatively, the EEG from each channel, regardless of whether or not there is a SSW on that particular channel, could be added to all previous sections of EEG on the same channel. Such processing may highlight otherwise hidden interchannel relationships that occur in the EEG in the vicinity of epileptiform

events.

Finally, it is important to note that a system for automatic detection of epileptiform activity, such as that presented here, applied in conjunction with other techniques for automated EEG analysis (Ktonas 1983; Gevins 1984), would comprise an extremely powerful and useful automated EEG analysis system.

Appendix A

The Electroencephalogram

Electroencephalography is one medical field in which information that is important for a patient's clinical diagnosis can be obtained through signal analysis. This field is briefly discussed here in order that the non-medically trained reader is prepared for the description of a new signal processing algorithm that is presented in Appendix B. This algorithm is designed to implement one facit of electroencephalograpy — the detection of activity characteristic of epilepsy. Consequently, only concepts of electroencephalography relevant to this task are discussed here.

In electroencephalography the signal that is analysed is the patient's *electroencephalogram* (EEG). The source of the EEG are the electrical potentials generated by the nerve cells in the brain. These fluctuating potentials summate and penetrate the scalp where they can be recorded as the scalp EEG. Specially designed electrodes are usually placed upon the scalp to record the potentials[1] (Chatrian *et al.* 1974). Typically 16 electrodes are distributed over the top of the head, with additional 'reference' electrodes being positioned at other locations on the body (see Fig. A.1). The electrodes are connected in certain combinations or *montages.*

Electrical signals in the form of *ionic currents* on the scalp originate primarily from activity in those parts of the brain closest to the position of signal measurement. Consequently, an EEG is primarily a measure of the electrical activity in the outer layers of the brain (see Fig. A.2).

Electrical signals that originate in the brain must pass through the skull before reaching the scalp. This causes the signal to be greatly attenuated, and also diffused, so that a signal of only approximately 50 – 100 μV is recorded by the electrodes (Spehlmann 1981). Consequently, these signals must be amplified before they can be usefully processed.

The *distribution* of the EEG refers to the occurrence of electrical activity recorded by electrodes positioned over different parts of the head. *Widespread* or *generalized* distribution refers to activity that occurs at the same time over most or all of the head. *Focal* activity describes that which is restricted to one or a few electrodes over an area of the head.

Considerable artefact (Chatrian *et al.* 1974) typically appears in scalp recorded EEGs due to other unwanted electrical signals impinging upon the electrodes. This extraneous signal invariably complicates the task of analysing an EEG, with the algorithm presented in Appendix B being no exception. Common causes of artefact are summarized here:

[1] Alternative methods for recording the electrical activity of the brain, usually involving the surgical implantation of electrodes under the skull, have been developed. These are usually only used during surgery, and are not necessarily for routine EEG recording. A recording made by such implanted electrodes is called an electrocortiogram (ECoG) (Chatrian et al. 1974).

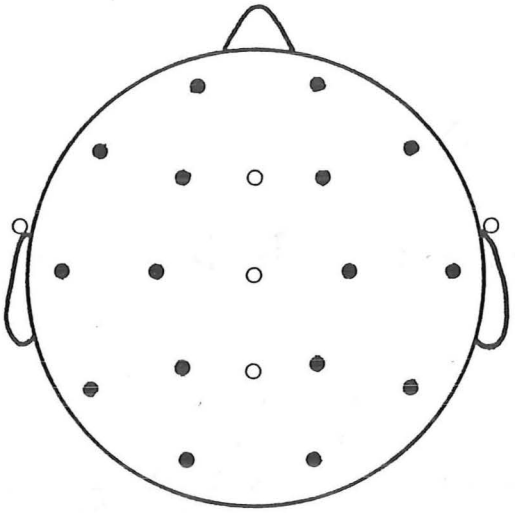


Figure A.1: A typical positioning of electrodes on the human head for recording EEG. The principal 16 electrodes (represented by filled circles) are placed on the scalp, with additional reference electrodes placed elsewhere (on the ears for example).

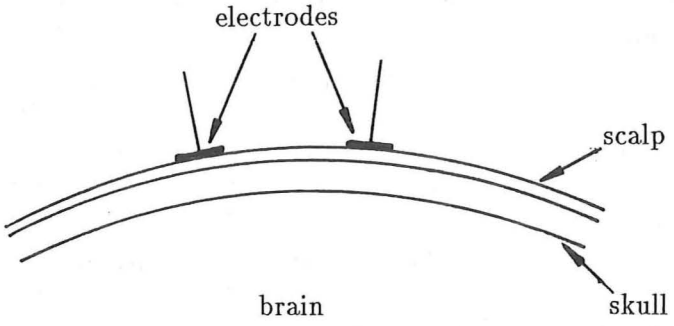


Figure A.2: Schematic diagram showing the relative position of brain, skull, scalp and electrodes. The brain tissue generates an electrical signal that is recorded by the electrodes. Note that the signal originating deep in the cerebrum is usually very highly attenuated when it reaches the electrodes.

(i) Muscle tissue on the scalp, face and jaw produces an unwanted electrical signal which appears in the EEG, because it impinges upon the recording electrodes. This *muscle artefact* attains a disproportionately large amplitude in the recorded EEG since it originates in the muscle tissue external to the skull, and is therefore not attenuated by the passage through the bone.

(ii) When a patient's eyes open or close artefact is often detected in the EEG. Opening the eyes tends to temporarily raise the recorded voltage, whereas blinking or closing the eyes tends to depress the voltage. This *eye-blink artefact* is detected predominantly on the frontal electrodes and the voltage fluctuations typically last for a quarter to one third of a second.

(iii) *Electrode artefact*, due to the electrodes placed on the scalp not making good electrical contact with the skin also commonly arises in the EEG. Movement of the electrodes is also a common cause of this type of artefact.

(iv) Potential changes generated by the heart are sometimes also picked up by the EEG electrodes.

Special EEG recording procedures designed to minimize artefacts occurring during recording are routinely implemented (Spehlmann 1981). These procedures require that the patient lie relaxed with eyes closed for most of the recording session, which typically lasts for half an hour or so.

From reading the EEG a trained medical specialist (in this field an *electroencephalographer* (EEGer)), is able to glean information about the function of the brain, and often is able to diagnose and localize certain neurological disorders (Spehlmann 1981; Ktonas 1983). One disorder that can often be diagnosed from the electroencephalogram is *epilepsy*. It is known that certain features appearing in an EEG are consistent with a patient suffering from epilepsy and EEGers routinely read EEGs for evidence of this *epileptiform* activity. One type of waveform that is known to be epileptiform has a 'spiky' nature and appears superimposed upon the ongoing EEG waveform. These structures are called *spikes* or *sharp-waves*, depending upon their duration (see §B.2).

The presence of artefact in EEG is usually inevitable, but the significance of the artefact to the analysis of an EEG is dependent upon the type and amplitude of the artefact, as well as the particular analysis to be performed. Analysis of the EEG is usually hindered by the need to discriminate between that component of the signal which is the true EEG and that which is artefact. The analysis is further complicated by the diverse range of sources within the brain producing the EEG signal.

When reading EEGs, EEGers unconsciously employ intricate pattern matching techniques learnt from many months of training and years of experience. Meaningful automated analysis of the EEG requires a computer to apply similar sorts of techniques. Computers, as yet, do not have the remarkable *pattern recognition* capacity of the human brain. Consequently, there are considerable limitations on the effectiveness of automated analysis of the EEG. However, algorithms for automating the analysis of specific parts of the interpretation process have proven useful (for reviews see Ktonas 1983; Gevins 1984; Gotman 1985).

Such an algorithm is presented in Appendix B, which describes a new computer-based system for detecting the presence of epileptiform activity in the EEG. The aim of this research was to automated assistance to an EEGer when reading an EEG for epilepsy. Many attempts to automate this *spike detection* process have been made over the last 20 or so years (*cf.* Gotman 1985), usually with only limited

success. The algorithm presented in Appendix B, however, is the first to use expert system techniques for this task. Preliminary results obtained from this system clearly demonstrate its potential for detecting and classifying epileptiform activity.

Appendix B

An EEG Spike-Detection System[1]

Abstract

An expert system for the automated detection of spikes and sharp waves in the EEG has been developed. The system consists of two distinct stages. The first is a feature extractor, written in the conventional procedural language FORTRAN, which utilises parts of previously published spike-detection algorithms to produce a list of all spike-like occurrences in the EEG. The second stage, written in the production system language OPS5, reads the list and uses rules incorporating knowledge elicited from an electroencephalographer (EEGer) to confirm or exclude each of the possible spikes. Information such as the time of occurrence, polarity and channel relationship are used in this process. A summary of the detected epileptiform events is produced which is available to the EEGer in interpreting the EEG. The performance of the expert system is compared with an EEGer using a 320 second segment from an EEG containing epileptiform activity. The system detected 19 events, and missed 7 (false negative), which the EEGer considered epileptiform. There were no false positive detections.

B.1 Introduction

There is much to be gained by the introduction of some form of automation into the process of interpreting EEGs, particularly in terms of increased uniformity, reduction of time spent in reading the EEG and for long-term EEG monitoring. Consequently numerous attempts to automate the detection of epileptiform activity have been made over the past 20 or so years (for reviews see, for example, Ktonas 1983 and Gotman 1985). Very few of these methods have been introduced into routine clinical

[1]The contents of this appendix comprise the paper:
Expert system approach to detection of epileptiform activity in the EEG

B.L.K. Davey
Department of Electrical and Electronic Engineering, University of Canterbury
Christchurch, New Zealand.
W.R. Fright
Department of Medical Physics and Bioengineering, Christchurch Hospital
Christchurch, New Zealand.
G.J. Carroll
Department of Neurology, Christchurch Hospital
Christchurch, New Zealand.
R.D. Jones
Department of Medical Physics and Bioengineering, Christchurch Hospital
Christchurch, New Zealand.

which currently in press (Davey *et al.* 1989a).

use and even then only with limited success, the main reason being the difficulty in writing computer programs to replicate human performance (Gotman 1985).

Arakawa *et al.* (1986) and Glover *et al.* (1986; 1987) have indicated that expert system techniques may hold promise for the detection of epileptiform activity in the EEG. Expert system techniques have resulted from research into artificial intelligence (Hayes-Roth 1985) and provide a method of representing, storing and applying expert knowledge with a computer. These techniques are finding increasing application in medicine (Proc. IEEE/EMBS 1987).

We have developed an automated spike-detection system comprising two analysis stages which marks a departure from previous approaches. The first stage is a feature extractor which utilises parts of previously published spike-detection algorithms to detect all spike-like occurrences in the EEG. Parameters defining possible spikes are passed to the second stage, which is a particular type of expert system known as a production system. The expert knowledge in these systems is embodied in the form of rules (or productions). Our productions identify epileptiform events mainly on the basis of the spatio-temporal relationships between transients.

A preliminary evaluation of the system's performance has been made by comparison with an electroencephalographer (EEGer) using a 320 second segment of an EEG containing epileptiform activity.

B.2 Epileptiform Activity in the EEG

An EEG is a recording of the electrical activity of the outer layer of the cerebral cortex. It is obtained by means of electrodes placed on the scalp and is most useful in the investigation and management of patients with suspected epilepsy. The recording of electrocerebral activity during one of a patient's clinical attacks may be particularly helpful in determining whether the attacks are epileptic in nature. Since seizures usually occur infrequently and unpredictably, the chances of a recording actually being in progress during an attack are not good, unless very prolonged recordings are made or seizures are deliberately provoked. Therefore, relatively short recordings (20 to 30 minutes) between seizures (inter-ictal) are more commonly made, during which only occasional epileptiform activity may be recorded. An EEGer subsequently reads the EEG to detect, amongst other things, the presence of epileptiform activity, but this is both time consuming and subject to error. While the presence of epileptiform activity in the inter-ictal EEG of such a patient is consistent with a clinical diagnosis of epilepsy, it does not establish the diagnosis beyond doubt, since similar activity may occasionally be recorded in subjects who have never had a seizure. Conversely, a normal inter-ictal recording does not exclude a diagnosis of epilepsy. Therefore, false negative errors (failure to report true epileptiform events) and, more importantly, false positive errors (reporting background activity or artefacts as epileptiform activity) must be minimized to avoid mis-diagnosis.

Epileptiform activity in the EEG is characterized by distinctive transient waveforms, *spikes* and *sharp waves*, both of which may occur alone or associated with *slow waves*. A spike is arbitrarily defined as "a transient, clearly distinguished from background activity, with pointed peak at conventional paper speeds and a duration from 20 to under 70ms" (Chatrian *et al.* 1974). A sharp wave is defined similarly but with a duration of 70–200ms. Hereafter we abbreviate spikes and sharp waves collectively as SSWs. Transients occurring in more than one channel synchronously can be further classified as focal or non-focal, in reference to their spatial distribution. Focal transients diminish in amplitude, and may decrease in sharpness, with dis-

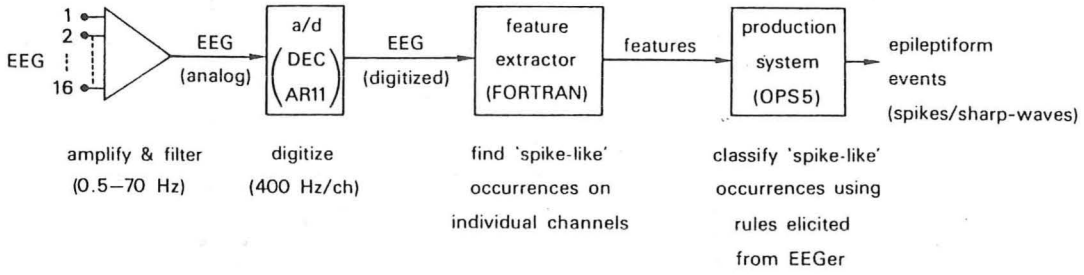


Figure B.1: Block diagram of the spike-detection system. Sampled EEG is processed by the feature extraction stage which produces a list of all *possible spikes*. This list is input to the production system which discriminates between the possible spikes to detect true epileptiform events. A report is output which is available to the EEGer to assist in the interpretation of the EEG.

tance from the focus. They can also exhibit *phase reversal* in bipolar recordings (see §B.3). Non-focal transients occur on several channels with similar sharpness, amplitude and polarity. Epileptiform transients cannot be simply defined and require years of accumulated experience to recognise reliably.

B.3 Data Acquisition

A block diagram of our system is presented in Fig. B.1. An EEG machine (Siemens 16 channel Minograph Universal) provides the amplification and an electrode configuration which is commonly used for EEG recording (10-20 system, see Fig. B.3(a)) to produce sixteen channels of EEG. Each channel is the voltage difference between an electrode and that immediately posterior. A focal event gives rise to positive transients in channels anterior to the focus and negative transients in channels posterior to the focus. This phenomenon is known as phase reversal and is very useful in localising focal events.

The EEG is low pass filtered (five pole, cut-off frequency 70 Hz) and digitised to ten bits at 400 Hz/channel by a PDP 11/34 computer. The data are stored for later processing on a VAX computer. While 200 Hz sampling would be adequate from an information theory viewpoint, closer spaced samples are necessary if pattern recognition algorithms are to be simple and reliable.

B.4 Feature Extraction

The raw data are processed by a feature extraction computer program, which detects spike-like occurrences in the EEG. The feature extractor is written in the procedural language FORTRAN because it is efficient at analysing numerical data. In principle,

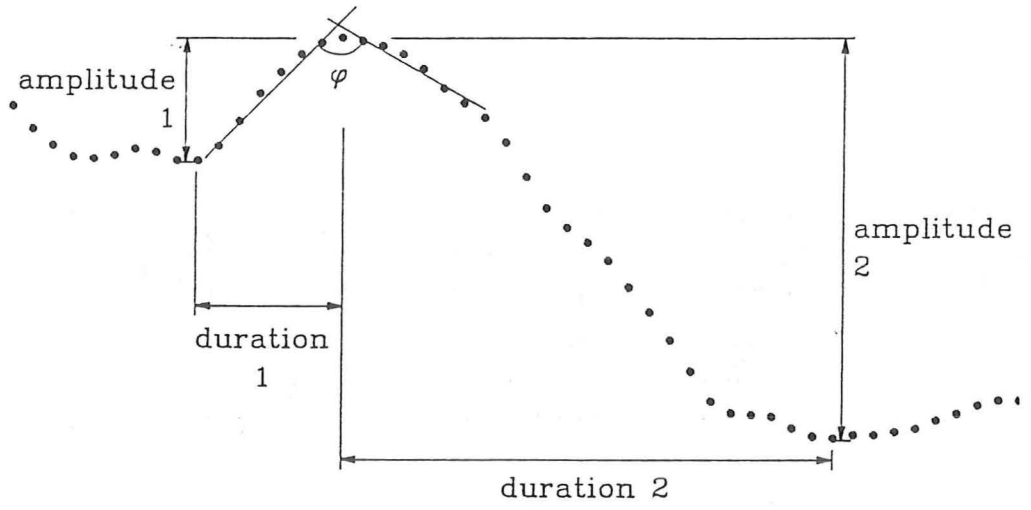


Figure B.2: Attributes of the waveform that are tested by the feature extractor when deciding if a particular waveform is possibly epileptiform.

the production system could process the raw EEG data directly, thereby omitting the feature extraction stage, but the execution time of the production system processing the raw data would be impractically long with currently available production systems.

Because the responsibility for minimizing false positive errors is mainly left to the production system stage, numerical parameters in the feature extraction stage are adjusted so that spike-like waveforms are detected with a minimal chance of false negative error. This results in far more possible SSWs being detected by the feature extractor than are eventually confirmed as true SSWs.

To isolate the possible SSWs, the feature extraction program operates in a series of steps. The raw EEG is first subdivided into a series of half-waves (*cf.* Gotman and Gloor 1976). We define a half-wave to be a segment of the EEG located between two adjacent extrema (i.e. a minimum and a maximum, or vice versa) and a wave to be a segment of EEG composed of two contiguous half-waves. The extrema are detected as a change in sign of the local slope, which we measure simply as the difference between samples spaced three apart. Taking every third sample, rather than contiguous samples, reduces the effect of small fluctuations due to noise and artefact superimposed upon the overall shape of the waveform.

Each wave is then tested to ascertain whether characteristics of the wave allow it to be classified as a possible SSW. The characteristics of the wave we test are duration, amplitude and peak sharpness. If any of these characteristics do not meet specified criteria, the current wave is disregarded and the processing progresses to the next wave. Ktonas *et al.* (1981) reported needing ten such characteristics to "quantify the morphology of an EEG spike". However, we are only locating SSW-like occurrences and have determined that the three characteristics are sufficient for the feature extraction stage.

The duration criteria are satisfied if the durations of the wave's component half-waves and the duration of the total wave are determined to fall between lower and upper bounds (see the first row of Table B.1). A 35ms lower bound for the total wave duration is chosen to reduce the number of false positive possible spikes produced as a result of EMG artefact, which resembles very short duration spikes (Gotman and Gloor 1976). A 200ms upper bound ensures that possible sharp waves are detected along with possible spikes.

parameter	larger half-wave		smaller half-wave		total wave	
	min.	max.	min.	max.	min.	max.
duration (ms)	15	120	15	120	35	200
amplitude factor	0.55	-	0.50	-	-	-
peak sharpness (φ)	-	-	-	-	-	16.0

Table B.1: Threshold parameters for the feature extraction stage.

If the absolute amplitude of both of the half-waves is less than a lower bound then the waveform is not a possible SSW. In the manner of Gotman and Gloor (1976) the amplitude bound is calculated as a fraction (we call it the background amplitude factor) of an estimate of the background amplitude of the EEG on that channel. We calculate the background amplitude of each channel as the average of the absolute amplitudes of the half-waves on that channel for the 5.12 seconds prior to the wave under consideration. We noticed that when SSWs are of low amplitude they often have one component half-wave of lower absolute amplitude than the other. Thus, two background amplitude factors are defined (Table B.1), the first for the half-wave with larger absolute amplitude and the second for the lower half-wave.

Peak sharpness of the wave is calculated by a procedure similar to Hill and Townsend (1973). Least squares regression is used to fit a straight line to six data points on the two constituent half-waves. The points chosen are those immediately adjacent to that point which forms the extremum (maximum or minimum) of the wave. The angle between the two fitted straight lines is calculated as an estimate of the peak angle and, if less than a set threshold (Table B.1), the waveform is considered a possible SSW.

A wave satisfying all the above criteria is a possible SSW and parameters defining the waveform are output to a file. The output parameters are

- the time of occurrence, channel, duration, peak angle and polarity of the wave
- the duration, amplitude and gradient of the two half-waves comprising the wave
- the estimate of the background amplitude on that channel.

B.5 Expert System

The second stage of processing is a production system which incorporates the detailed knowledge of an expert in order to discriminate between the possible SSWs. This stage is characterized by its multi-channel perspective and is implemented using the OPS5 production system language (Forgy 1981) run on a VAX computer.

B.5.1 Rule Based Production Systems

A production system is a program composed entirely of conditional statements called productions (or rules) (Hayes-Roth 1985; Brownston *et al.* 1985). A production is similar in construction to the IF-THEN statement of a conventional programming language. The general production containing m conditions $C_1 ... C_m$ and n actions $A_1 ... A_n$ can be written

IF conditions $C_1 \ldots C_m$ are true simultaneously
THEN execute the actions $A_1 \ldots A_n$.

The major difference between a production system and a conventional procedural language is the method of control over the program's execution. Because the data determine which conditions are satisfied, and hence which productions can fire, this type of production system is said to be *data-driven.* Use of production systems is appropriate when the knowledge occurs naturally in rule form.

Production systems typically have a three stage execution cycle. The first stage produces a list, called the conflict set, of all productions whose conditions are satisfied by data in memory. In the second stage, called conflict resolution, the production system uses ordering techniques to determine which of the productions in the conflict set is most relevant to execute. In the third stage the chosen production executes and usually modifies the data. The cycle is repeated until no production has all of its conditions satisfied, at which stage execution ceases.

B.5.2 Knowledge Elicitation

The motivation behind expert systems is to allow a computer to mimic the sequence of steps and reasoning used by an expert in solving a problem in their domain of expertise. Thus, the principal requirement in building an expert system is the availability of an expert from whom to obtain a strategy for solving the problem.

The knowledge applied by an EEGer in interpreting an EEG has been accumulated through many months of training and years of experience. Often the EEGer is not consciously aware of the detailed steps involved in his/her analysis of the recording, making elicitation of this expert knowledge difficult. The accurate acquisition of this knowledge is, however, vital as it forms the basis of the production system.

Discussions between electrical/biomedical engineers (BLKD, WRF, RDJ) and a clinical neurophysiologist (GJC) were held over several months to elicit the knowledge used by the neurophysiologist in detecting epileptiform activity in the EEG. We undertook a three stage iterative cycle of inquire, build and analyse. In each cycle a discussion was held, the new information arising from the discussion was encoded into the program, and the output of the system formed the basis for the next discussion. This allowed a deeper understanding of the problem and how an expert system may deal with it to be obtained by the engineers and also by the neurophysiologist.

B.5.3 Productions

The productions comprising our system can be split into two categories – those containing and applying the domain knowledge and those controlling the flow of data through the program.

The knowledge elicited from the neurophysiologist can be summarised in the following six categories:

Channel orientation – which pairs of channels are adjacent and the polarity relationship expected on each pair during focal activity. Currently only the electrode configuration depicted in Fig. B.3(a) is implemented. However, other configurations can easily be incorporated by defining a different set of channel adjacency relationships to specify that configuration. With reference to Fig. B.3(a), channel 1 is adjacent to channel 2 and, because the electrodes are *longitudinally* adjacent, a polarity reversal is expected for focal activity. In contrast, channels 1 and 5 are *transversally* adjacent and no polarity reversal is expected.

Synchrony – two SSWs that are separated in time by 10ms or less are related to the same event, otherwise they are considered to be due to separate events. It

is implicit in the remainder of the knowledge presented here that the mentioned SSWs relate synchronously to the same event.

Amplitude – an EEGer, when reading an EEG, is alerted to an epileptiform event if there is at least one SSW whose amplitude is significantly above that of the ongoing background activity. Thus, for the production system to detect an epileptiform event, at least one SSW must have an amplitude a certain factor (we use 0.7) above the background amplitude on its channel.

Duration – a SSW is a spike if its duration is less than 70ms, otherwise it is considered to be a sharp wave.

Focal event – occurs if there are two or more SSWs occurring on adjacent channels and the polarity relation for each pair of adjacent channels is the same as that expected for focal activity on these channels.

Non-focal event – occurs if there are SSWs occurring with negative polarity on three or more channels and the conditions for focal activity are not met.

An SSW arising on only one channel does not constitute an epileptiform event because both focal and non-focal events require SSWs on at least two channels.

All possible SSWs which are classified as being a component of an epileptiform event are tagged with the type of activity (i.e.either focal or non-focal) for further processing. Eight productions were required to implement this domain knowledge.

Once all possible SSWs have been processed the production system generates a report showing the number of spikes and sharp waves (separated into focal and non-focal categories) detected on each channel. A chronological list of the time of occurrence of epileptiform events is also output, detailing whether the event was focal or non-focal and all the channels on which the event occurred. Seven productions are used to fulfil this report generation function.

In addition to productions incorporating domain knowledge, productions are necessary for program control. Productions are needed to read new possible SSWs from the file and others to remove possible SSWs which have been deemed not to be true spikes or sharp waves. There are nine control productions which, although not applying any domain knowledge, are essential for the successful operation of the system.

B.6 Results

A preliminary evaluation of our system was provided by its performance on a 320 second segment of a routine EEG recorded from a 13 year old female who had a first grand mal epileptic seizure the previous day. The recording contained epileptiform activity as well as containing EMG, electrode and movement artefacts. The EEG was read by the neurophysiologist and also processed by the spike-detection system. Fig. B.3(b) shows a typical 7.8ms segment of the recording.

The feature extractor detected 545 possible SSWs whose parameters were passed to the production system. The production system classified 19 events as epileptiform, 17 of which had been marked as true epileptiform events by the neurophysiologist. The neurophysiologist, upon further examination of the recording, concluded that the remaining two events were, without doubt, epileptiform. The first was obscured because it occurred exactly at a page boundary and the second because it occurred during a segment of recording containing a large amount of artefacts. Thus, no false

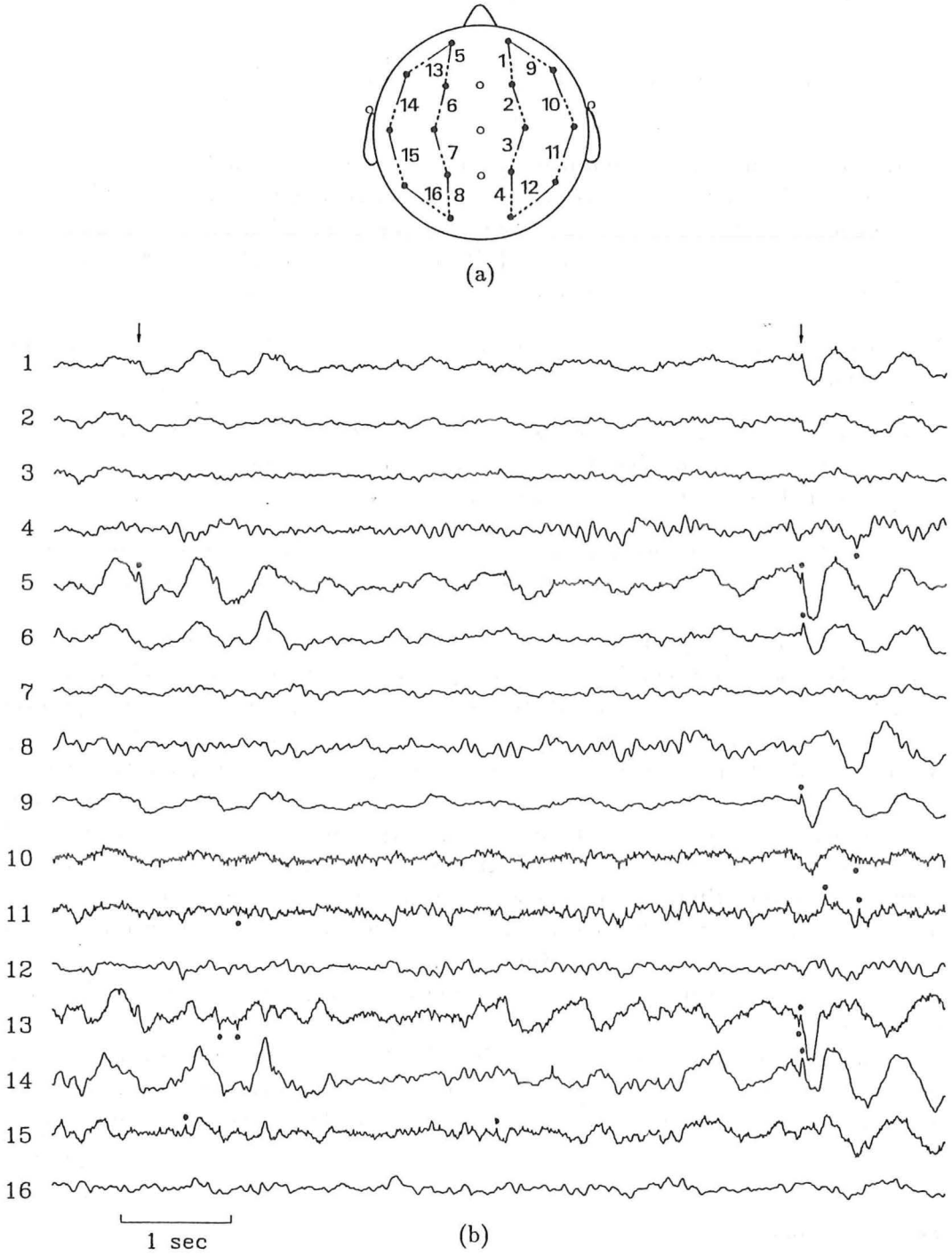


Figure B.3: (a) Electrode configuration used to record the EEG. (b) 7.8 seconds of the 320s segment of EEG used to develop and test the system. Possible SSWs detected by the feature extraction stage are indicated with dots. The first arrow (above channel 1) identifies a typical false negative event. The sharp wave on channel 5, labelled by the neurophysiologist as being epileptiform, was not detected by the production system stage because there was no supporting evidence on other channels to confirm the event (the transient on channel 13 was not sufficiently sharp to be considered epileptiform by the feature extractor and it had not been classified as epileptiform by the neurophysiologist). The second arrow identifies an event which the system correctly detected as being epileptiform. The SSWs that were used by the system in reaching this decision occur on channels 5, 6, 9 and 13. The SSW on channel 14 and the downward one on channel 13 are not part of this event.

positive events were detected. Of the 19 events, 17 were non-focal and occurred with a similar distribution while the other two arose focally on two other channels. This was detected by the production system and confirmed by the neurophysiologist.

The neurophysiologist marked six SSWs that were not detected by the system, two being spikes and four being sharp waves. The feature extraction stage detected each of these but, because they arose on only one channel, there was no supporting evidence on other channels and they were rejected as epileptiform events by the production system. One non-focal event was also marked but the feature extractor detected SSWs on only two channels and therefore the production system rejected this event.

B.7 Discussion

We have presented a system comprising a feature extractor and a production system for detecting epileptiform activity in the EEG. The system is also able to differentiate between focal and non-focal events. Preliminary results obtained from running this system on a 320s segment of EEG containing epileptiform activity, as well as artefacts, have demonstrated its potential for detecting and classifying epileptiform activity. The feature extractor correctly detected 19 epileptiform events and correctly classified each as being focal or non-focal. The circumstances regarding the non-detection of two events by the neurophysiologist reinforce the desirability of automated assistance in the analysis of EEGs. Of the seven false negative detections, six resulted because the event occurred on only one channel. The emphasis of our system is on reducing false positive events at the expense of allowing some false negative detections.

When reading an EEG for epileptiform activity an EEGer is more interested in the occurrence of epileptiform events than in individual SSWs. Therefore our results are presented as detected epileptiform events rather than the number of individual SSWs. The channels on which these SSWs occur are important in locating the focus of the abnormality, so the output of the production system also provides a summary of these channels.

Our system will be used to routinely aid one EEGer, the knowledge incorporated in the rules for the production system being elicited from that person. Other EEGers may suggest modifications to incorporate aspects of their expertise in interpreting EEGs. The production system has a major advantage over non-expert system approaches in being easy to modify by adding productions which incorporate new information.

For development and verification of the system, the raw data were transferred from a PDP computer to a VAX computer for further processing. For routine use we are implementing the data acquisition, feature extraction and production system stages on a personal computer. We recognise that the system evaluation presented here is limited, being based on only one patient sample. A clinical study involving a number of patients is to be undertaken and reported in the future.

An expert system such as that reported here for spike-detection combined with automated analysis of the background EEG would provide for extremely powerful EEG analysis.

References

ABBOTT, J.G. and THURSTONE, F.L. (1979), 'Acoustic speckle: Theory and experimental analysis', *Ultrasonic Imaging*, Vol. 1, No. 4, October, Pp. 303–324.

AIME, C., RICORT, G., RODDIER, C. and LAGO, G. (1978), 'Changes in the atmospheric-lens modulation transfer function used for calibration in solar speckle interferometry', *Journal of the Optical Society of America A*, Vol. 68, No. 8, August, Pp. 1063–1066.

AITKEN, G.J.M., HOUTMAN, R., JOHNSON, R. and POCHET, J. (1985), 'Direct phase gradient measurement for speckle image reconstruction', *Applied Optics*, Vol. 24, No. 18, 15 September, Pp. 2926–2930.

AITKEN, G.J.M., JOHNSON, R. and HOUTMAN, R. (1986), 'Phase-gradient stellar image reconstruction', *Optics Communications*, Vol. 56, No. 6, 15 January, Pp. 379–383.

ANDREWS, H.C. and HUNT, B.R. (1977), *Digital Image Restoration*, Prentice-Hall, New Jersey.

ARAKAWA, K., FENDER, D., HARASHIMA, H., MIYAKAWA, H. and SAITOH, Y. (1986), 'Separation of a nonstationary component from the EEG by a nonlinear digital filter', *IEEE Transactions on Biomedical Engineering*, Vol. BME-33, No. 7, July, Pp. 724–726.

ARRATHOON, R. (Ed.) (1986), *Digital Optical Computing*, Vol. 25, Optical Engineering.

ARSENAULT, H.H. and CHALASINSKA-MACUKOW, K. (1983), 'The solution to the phase retrieval problem using the sampling theorem', *Optics Communications*, Vol. 47, No. 6, October, Pp. 380–386.

AYERS, G.R. and DAINTY, J.C. (1988), 'An iterative blind deconvolution algorithm and its applications', *Optics Letters*, Vol. 13, No. 7, July, Pp. 547–549.

AYERS, G.R., NORTHCOTT, M.J. and DAINTY, J.C. (1988), 'Knox-Thompson and triple-correlation imaging through atmospheric turbulence', *Journal of the Optical Society of America A*, Vol. 5, No. 7, July, Pp. 963–985.

BABA, N., KAWAGUCHI, F., OSE, T. and ISOBE, S. (1984), 'Phase unwrapping method along radial coordinates for speckle image reconstruction', *Optics Communications*, Vol. 49, No. 1, 1 February, Pp. 11–16.

BABA, N., ISOBE, S., NORIMOTO, Y. and NOGUCHI, M. (1985), 'Stellar speckle image reconstruction by the shift-and-add method', *Applied Optics*, Vol. 24, No. 10, 15 May, Pp. 1403–1405.

BABA, N., ISOBE, S., NOGUCHI, M., NORIMOTO, Y. and MIURA, N. (1987), 'Data reduction for stellar speckle interferograms of binary stars based on the shift-and-add method', *Applied Optics*, Vol. 26, No. 12, 15 June, Pp. 2306–2310.

BAGNUOLO, JR., W.G. (1982), 'The application of Bates' algorithm to binary stars', *Monthly Notices of the Royal Astronomical Society*, Vol. 200, No. 3, September, Pp. 1113–1122.

BAGNUOLO, JR., W.G. (1985), 'Image restoration by the shift-and-add algorithm', *Optics Letters*, Vol. 10, No. 5, May, Pp. 200–202.

BAGNUOLO, JR., W.G. (1988), 'Binary-star intensity ratios by the fork algorithm', *Optics Letters*, Vol. 13, No. 10, October, Pp. 907–909.

BAGNUOLO, JR., W.G. and SOWELL, J.R. (1988), 'Binary star speckle photometry. I. the colours and spectral types of the Capella stars', *The Astronomical Journal*, Vol. ? Submitted 29 Feb 1988.

BARLOW (1975), *The Astronomical Telescope*, Wykeham Publications, London.

BARTELT, H., LOHMANN, A.W. and WIRNITZER, B. (1984), 'Phase and amplitude recovery from bispectra', *Applied Optics*, Vol. 23, No. 18, 15 September, Pp. 3121–3129.

BATCHELOR, B.G. (Ed.) (1978), *Pattern Recognition: Ideas In Practice*, Plenum Press, New York and London.

BATES, R.H.T. (1969), 'Contributions to the theory of intensity interferometry', *Monthly Notices of the Royal Astronomical Society*, Vol. 142, No. 4, Pp. 413–428.

BATES, R.H.T. (1976), 'A stochastic image restoration procedure', *Optics Communications*, Vol. 19, No. 2, November, Pp. 240–244.

BATES, R.H.T. (1978), 'On phase problems. II', *Optik*, Vol. 51, No. 3, September, Pp. 223–234.

BATES, R.H.T. (1982a), 'Astronomical speckle imaging', *Physics Reports*, Vol. 90, Pp. 203–297.

BATES, R.H.T. (1982b), 'Fourier phase problems are uniquely solvable in more than one dimension. I: Underlying theory', *Optik*, Vol. 61, No. 3, June, Pp. 247–262.

BATES, R.H.T. (1987a), 'Aspects of the Erlangen-bispectrum', *Optik*, Vol. 76, Pp. 23–26.

BATES, R.H.T. (1987b), 'Some image processing: Highlights in retrospect', *Search*, Vol. 18, No. 5, September, Pp. 237–240.

BATES, R.H.T. and CADY, F.M. (1980), 'Towards true imaging by wideband speckle interferometry', *Optics Communications*, Vol. 32, No. 3, March, Pp. 365–369.

BATES, R.H.T. and DAVEY, B.L.K. (1987a), 'Computationally-cost-effective speckle imaging', In *Proceedings of the Joint Workshop on High-Resolution Imaging from the Ground Using Interferometric Techniques*, Oracle, Arizona, January 12–15, Pp. 219–222.

BATES, R.H.T. and DAVEY, B.L.K. (1987b), 'Towards making shift-and-add a versatile imaging technique', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol.* 828*: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 87–94.

BATES, R.H.T. and DAVEY, B.L.K. (1988), 'Deconvolution ancient and (very) modern', In ALLOIN, D. and MARIOTTI, J.M. (Eds.), *Proceedings of the Cargèse Advanced Study Institute Conference on Diffraction Limited Imaging with Very Large Telescopes*, NATO, Kluwer Academic Publishers.

BATES, R.H.T. and FRIGHT, W.R. (1982), 'Towards imaging with a speckle-interferometric optical synthesis telescope', *Monthly Notices of the Royal Astronomical Society*, Vol. 198, No. 3, Pp. 1017–1031.

BATES, R.H.T. and LANE, R.G. (1987a), 'Deblurring should now be automatic', In *6th Pfefferkorn Conference on Image and Signal Processing in Electron Microscopy, Niagara Falls, Ontario, Canada*, April.

BATES, R.H.T. and LANE, R.G. (1987b), 'Automatic deconvolution and phase retrieval', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol.* 828*: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 158–164.

BATES, R.H.T. and LANE, R.G. (1988), 'Deblurring should now be automatic', In *Scanning Microscopy Supplement 2*, AMF O'Hare, Chicago, Pp. 149–156.

BATES, R.H.T. and M^{c}DONNELL, M.J. (1986), *Image Restoration and Reconstruction*, Clarendon Press, Oxford.

BATES, R.H.T. and MILNER, M.O. (1978), 'Towards imaging of star clusters by speckle interferometry', In VAN SCHOONEVELD, C. (Ed.), *Proceedings of IAU Colloquim No. 75: Image Formation from Coherence Functions in Astronomy*, IAU, D. Reidal Publishing Company, Groningen, Netherlands, 10–12 August, Pp. 187–193.

BATES, R.H.T. and MINARD, R.A. (1984), 'Compensation for multiple reflection', *IEEE Transactions on Sonics and Ultrasonics*, Vol. SU-31, No. 4, July, Pp. 330–336.

BATES, R.H.T. and MNYAMA, D. (1986), 'The status of practical Fourier phase retrieval', In HAWKES, P.W. (Ed.), *Advances in Electronics and Electron Physics*, Academic, Pp. 1–64.

BATES, R.H.T. and NAPIER, P.J. (1972), 'Identification and removal of phase errors in interferometry', *Monthly Notices of the Royal Astronomical Society*, Vol. 158, No. 4, Pp. 405–424.

BATES, R.H.T. and ROBINSON, B.S. (1982), 'A stochastic imaging procedure', In ASH, E.A. and HILL, C.R. (Eds.), *Acoustical Imaging*, Plenum Press, New York, Pp. 185–191.

BATES, R.H.T., NAPIER, P.J., McKINNON, A.E. and McDONNELL, M.J. (1976), 'Self consistent deconvolution. I: Theory', *Optik*, Vol. 44, No. 2, January, Pp. 183–201.

BATES, J.H.T., McKINNON, A.E. and BATES, R.H.T. (1982), 'Subtractive image restoration. II: Comparison with multiplicative deconvolution', *Optik*, Vol. 62.

BATES, R.H.T., GARDEN, K.L. and PETERS, T.M. (1983), 'An overview of computerized tomography with emphasis on future developments', *Proceedings IEEE*, Vol. 71, No. 3, March, Pp. 356–372. Invited paper for the Special Issue on Computed Tomography.

BATES, J.H.T., FRIGHT, W.R. and BATES, R.H.T. (1984), 'Wiener filtering and cleaning in a general image processing context', *Monthly Notices of the Royal Astronomical Society*, Vol. 211, No. 1, 1 November, Pp. 1–14.

BATES, R.H.T., SINTON, A.M. and MINARD, R.A. (1985), 'Generalization of shift-and-add imaging', In ARSENAULT, H.H. (Ed.), *Proceedings SPIE: Vol. 556: International Conference on Speckle*, SPIE, August 20-23, Pp. 263–269.

BATES, R.H.T., BRIESEMAN, N.P., CLARK, T.M., ELDER, A.G., FRIGHT, W.R., GARDEN, K.L., KENNEDY, W.K., SQUIRES, P.L., THORPE, C.W., JELINEK, H.J. and TURNER, S.G. (1987a), 'Interactive speech-defect diagnostic/therapeutic/prosthetic aid', In *Proceedings SPIE: Vol. 827: Real Time Signal Processing X*, SPIE, Pp. 131–136.

BATES, R.H.T., FRIGHT, W.R. and GARDENIER, P.H. (1987b), 'Gerchberg-Saxton phase retrieval when image magnitude given only approximately', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol. 828: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 87–94.

BLAKE, L.V. (1984), *Antennas*, Artech House Inc., Norwood, Massachusetts, 2 ed.

BORN, M. and WOLF, E. (1970), *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Pergamon Press, Oxford, 5 ed.

BRACEWELL, R.N. (1978), *The Fourier Transform and its Applications*, McGraw-Hill, New York.

BRAMES, B.J. (1986), 'Unique phase retrieval with explicit support information', *Optics Letters*, Vol. 11, Pp. 61–63.

BRAMES, B.J. (1987), 'Efficient method of support reduction', *Optics Communications*, Vol. 64, No. 4, 15 November, Pp. 333–337.

BRIESEMAN, N.P., THORPE, C.W. and BATES, R.H.T. (1987), 'Nontactile estimation of glottal excitation characteristics of voiced speech', *Proceedings IEE, Part A*, Vol. 134, No. 10, December, Pp. 807–813.

BROWNSTON, L., FARRELL, R., KANT, E. and MARTIN, N. (1985), *Programming expert systems in OPS5: An introduction to rule-based programming*, Addison-Wesley, Reading, Massachusetts.

BRUCK, Y.M. and SODIN, L.G. (1979), 'On the ambiguity of the image reconstruction problem', *Optics Communications*, Vol. 30, No. 3, September, Pp. 304–308.

BRUCK, Y.M. and SODIN, L.G. (1983), 'An improved method for reconstruction of two- and multidimensional images from the phase of their Fourier spectrum', *Optica Acta*, Vol. 30, Pp. 995–999.

BRUCK, Y.M. and SODIN, L.G. (1984), 'Speckle interferometry image reconstruction from the Fourier transform phase', *Journal of the Optical Society of America A*, Vol. 1, No. 1, January, Pp. 73–80.

BRYAN, R.K. and SKILLING, J. (1986), 'Maximum entropy image reconstruction from phaseless Fourier data', *Optica Acta*, Vol. 33, Pp. 287–299.

BURGE, R.E., FIDDY, M.A., GREENAWAY, A.H. and ROSS, G. (1976), 'The phase problem', *Proceedings Royal Society of London A*, Vol. 350, Pp. 191–212.

CADY, F.M. (1980), *Applications of Microcomputers in Interactive Image Processing*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

CEDERQUIST, J.N., FIENUP, J.R., MARRON, J.C. and PAXMAN, R.G. (1988), 'Phase retrieval from experimental far-field speckle data', *Optics Letters*, Vol. 13, No. 8, August, Pp. 619–621.

CHATRIAN, G.E., BERGAMINI, L., DONDEY, M., KLASS, D.W., LENNOX-BUCHTHAL, M. and PETERSÉN, I. (1974), 'A glossary of terms most commonly used by clinical electroencephalographers', *Electroencephalography and Clinical Neurophysiology*, Vol. 37, Pp. 538–548.

CHELLI, A., LÉNA, P., RODDIER, C., RODDIER, F. and SIBILLE, F. (1979), 'Modulation transfer function for infra-red stellar speckle interferometry: Evidence for a log-normal statistic', *Optica Acta*, Vol. 26, No. 5, May, Pp. 583–595.

CHRISTIANSEN, W.N. and HÖGBOM, J.A. (1969), *Radiotelescopes*, Cambridge University Press, Cambridge, UK, 2 ed.

CHRISTOU, J.C., HEGE, E.K., FREEMAN, J.D. and RIBAK, E. (1985), 'Images from astronomical speckle data: weighted shift-and-add analysis', In ARSENAULT, H.H. (Ed.), *Proceedings SPIE: Vol. 556: International Conference on Speckle*, SPIE, 20–23 August, Pp. 255–262.

CHRISTOU, J.C., HEGE, E.K., FREEMAN, J.D. and RIBAK, E. (1986a), 'Self-calibrating shift-and-add technique for speckle imaging', *Journal of the Optical Society of America A*, Vol. 3, No. 2, February, Pp. 204–209.

CHRISTOU, J.C., RIBAK, E., HEGE, E.K. and FREEMAN, J.D. (1986b), 'Images from astronomical speckle data: weighted shift-and-add analysis', *Optical Engineering*, Vol. 25, No. 6, June, Pp. 724–730.

CHRISTOU, J.C., M^{c}CARTHY, D.W. and COBB, M.L. (1987a), 'Image selection and binning for improved atmospheric calibration of infrared speckle data', *The Astronomical Journal*, Vol. 94, No. 2, August, Pp. 516–522.

CHRISTOU, J.C., FREEMAN, J.D., RODDIER, F., M^{c}CARTHY, D.W., COBB, M.L. and SHAKLAN, S.B. (1987b), 'Application of bispectrum analysis to infrared speckle data', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol. 828: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 32–39.

COLE, T.W. (1977), 'Quasi-optical techniques of radio astronomy', In WOLF, E. (Ed.), *Progress in Optics*, North-Holland, Amsterdam, Chap. IV, Pp. 187–244.

COOLEY, J.W. and TUKEY, J.W. (1965), 'An algorithm for the machine calculation of complex Fourier series', *Mathematics of Computation*, Vol. 19, No. 90, April, Pp. 297–301.

CRIMMINS, T.R. and FIENUP, J.R. (1983), 'Uniqueness of phase retrieval for functions with sufficiently disconnected support', *Journal of the Optical Society of America*, Vol. 73, Pp. 218–221.

CURTIS, S.R. and OPPENHEIM, A.V. (1987), 'Reconstruction of multidimensional signals from zero crossings', *Journal of the Optical Society of America A*, Vol. 4, Pp. 221–231.

CURTIS, S.R., OPPENHEIM, A.V. and LIM, J.S. (1985), 'Signal reconstruction from Fourier transform sign information', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-33, Pp. 643–657.

DAINTY, J.C. (1974), 'The transfer function, signal-to-noise-ratio, and limiting magnitude in stellar speckle interferometry', *Monthly Notices of the Royal Astronomical Society*, Vol. 169, No. 3, December, Pp. 631–641.

DAINTY, J.C. (1978), 'Computer simulations of speckle interferometry of binary stars in the photon counting mode', *Monthly Notices of the Royal Astronomical Society*, Vol. 183, Pp. 223–236.

DAINTY, J.C. (1984), 'Stellar speckle interferometry', In DAINTY, J.C. (Ed.), *Laser Speckle and Related Phenomena*, Springer-Verlag, Berlin, Chap. 7, Pp. 255–320.

DAINTY, J.C. and NORTHCOTT, M.J. (1986), 'Imaging a randomly translating object at low light levels using the triple correlation', *Optics Communications*, Vol. 58, No. 1, 1 May, Pp. 11–14.

DAVEY, B.L.K. and THORPE, C.W. (1987), 'Image and signal reconstruction by shift-and-add', In *IPENZ conference proceedings, Vol. 2*, IPENZ, Christchurch, New Zealand, May, Pp. 147–157.

DAVEY, B.L.K., SINTON, A.M. and BATES, R.H.T. (1986), 'Zero-and-add', *Optical Engineering*, Vol. 25, No. 6, June, Pp. 765–771.

DAVEY, B.L.K., LANE, R.G. and BATES, R.H.T. (1988a), 'Blind deconvolution of noisy complex-valued image', *Optics Communications*, Vol. 69, No. 5–6, 15 January, Pp. 353–356.

DAVEY, B.L.K., FRIGHT, W.R., CARROLL, G.J. and JONES, R.D. (1988b), 'An expert system for detecting epileptiform activity in the EEG', *New Zealand Medical Journal*, Vol. 101, 28 September, P. 610. Abstract of presentation for the 6 April 1988 meeting of the Christchurch Medical Research Society.

DAVEY, B.L.K., FRIGHT, W.R., CARROLL, G.J. and JONES, R.D. (1989a), 'An expert system approach to detection of epileptiform activity in the EEG', *Medical and Biological Engineering and Computing*, Vol. 27. To appear.

DAVEY, B.L.K., COCKE, W.J., BATES, R.H.T., M^{c}CARTHY, JR., D.W., CHRISTOU, J.C. and COBB, M.L. (1989b), 'Infrared speckle observations of binary Ross 614 AB: Combined shift-and-add and zero-and-add analysis', *The Astronomical Journal*, Vol. ? To appear.

DE FREITAS, L.C. and DAINTY, J.C. (1988a), 'Object reconstruction from photon-limited centroided data of randomly translating images', *Optics Letters*, Vol. 13, April, Pp. 264–266.

DE FREITAS, L.C. and DAINTY, J.C. (1988b), 'Centroiding as a phase closure relationship', In *Proceedings of NOAO/ESO Conference on High Resolution Imaging in Interferometry*, NOAO/ESO, Garching, Federal Republic of Germany, 15–18 March.

DE FREITAS, L.C., NORTHCOTT, M., BRAMES, B.J. and DAINTY, J.C. (1987), 'Object reconstruction from photon-limited centroided data of randomly translating images', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol. 828: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 62–73.

DEIGHTON, H.V., SCIVIER, M.S. and FIDDY, M.A. (1985), 'Solution of the two-dimensional phase retrieval problem', *Optics Letters*, Vol. 10, Pp. 250–251.

DINGLE, A.A. (1988), 'The processing and modelling of electroencephalograms', Third Professional Year Project Report, 23 September.

DRUMMOND, J.D., COCKE, W.J., HEGE, E.K., STRITTMATTER, P.A. and LAMBERT, J.V. (1985), 'Speckle interferometry of asteroids. I. 433 Eros', *Icarus*, Vol. 61, No. 1, January, Pp. 132–151.

DYCK, H.M. and HOWELL, R.R. (1985), 'Astronomical speckle imaging in the infrared', In ARSENAULT, H.H. (Ed.), *Proceedings SPIE: Vol.* 556*: International Conference on Speckle*, SPIE, 20-23 August, Pp. 274–278.

DYCK, H.M., ZUCERMAN, B., LEINERT, C. and BECKWITH, S. (1984), 'Near-infrared speckle interferometry of evolved stars and bipolar nebulae', *The Astrophysical Journal*, Vol. 287, December 15, Pp. 801–813.

EBERSBERGER, J. and WEIGELT, G.P. (1985), 'Isoplanacity and lifetime of stellar speckle interferograms', *Optica Acta*, Vol. 32, No. 7, July, Pp. 793–801.

EICHEL, P.H., GHIGLIA, D.C. and JAKOWATZ, JR., C.V. (1989), 'Speckle processing method for synthetic-aperture-radar phase correction', *Optics Letters*, Vol. 14, No. 1, January, Pp. 1–3.

FIDDY, M.A., BRAMES, B.J. and DAINTY, J.C. (1983), 'Enforcing irreducibility for phase retrieval in two dimensions', *Optics Letters*, Vol. 8, Pp. 96–98.

FIENUP, J.R. (1978), 'Reconstruction of an object from the modulus of its Fourier transform', *Optics Letters*, Vol. 3, No. 1, July, Pp. 27–29.

FIENUP, J.R. (1982), 'Phase retrieval algorithms: A comparison', *Applied Optics*, Vol. 21, No. 15, 1 August, Pp. 2758–2769.

FIENUP, J.R. (1983), 'Reconstruction of objects having latent reference points', *Journal of the Optical Society of America*, Vol. 73, No. 11, November, Pp. 1421–1426.

FIENUP, J.R. (1984), 'Comparison of phase retrieval algorithms', In HUANG, T.S. (Ed.), *Advances in Computer Vision and Image Processing: Vol. 1 — Image Reconstruction from Incomplete Observations*, JAI Press Inc., Greenwich, Connecticut, Chap. 4, Pp. 191–225.

FIENUP, J.R. (1987), 'Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint', *Journal of the Optical Society of America A*, Vol. 4, No. 1, January, Pp. 118–123.

FIENUP, J.R. and WACKERMAN, C.C. (1986), 'Phase-retrieval stagnation problems and solutions', *Journal of the Optical Society of America A*, Vol. 3, No. 11, November, Pp. 1897–1907.

FONTANELLA, J.C. and SÈVE, A. (1987), 'Reconstruction of turbulence-degraded images using the Knox-Thompson algorithm', *Journal of the Optical Society of America A*, Vol. 4, No. 3, March, Pp. 438–448.

FORGY, C.L. (1981), *OPS5 User's Manual*, Technical report CMU-CS-81-135, Department of Computer Science, Carnegie-Mellon University.

FREEMAN, J.D., RIBAK, E., CHRISTOU, J.C. and HEGE, E.K. (1985), 'Statistical analysis of the weighted shift-and-add reconstruction technique', In ARSENAULT, H.H. (Ed.), *Proceedings SPIE: Vol.* 556*: International Conference on Speckle*, SPIE, 20–23 August, Pp. 279–283.

FREEMAN, J.D., CHRISTOU, J.C., M^{c}CARTHY, D.W. and COBB, M.L. (1987), 'A comparison of phase retrieval algorithms applied to infrared astronomical speckle data', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol.* 828*: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 40–46.

FREEMAN, J.D., CHRISTOU, J.C., RODDIER, F., McCARTHY, JR., D.W. and COBB, M.L. (1988), 'Application of bispectrum analysis for phase recovery from one-dimensional infrared speckle data', *Journal of the Optical Society of America A*, Vol. 5, No. 3, March, Pp. 406–415.

FREIBERGER, W.F. (Ed.) (1960), *The International Dictionary of Applied Mathematics*, D. Van Nostrand Company, Inc., Princeton, New Jersey.

FRIED, D.L. (1966), 'Optical resolution through a randomly inhomogeneous medium for very long and very short exposures', *Journal of the Optical Society of America*, Vol. 56, No. 10, October, Pp. 1372–1379.

FRIGHT, W.R. (1984), *The Fourier Phase Problem*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

FRIGHT, W.R., BATES, R.H.T., LOWE, J.M.R., SPENCER, D.S., SOMEKH, M.G. and BRIGGS, G.A.D. (1988), 'Reconstruction of the complex reflectance function in acoustic microscopy', *Journal of Microscopy*. In press.

FROST, R.L., RUSHFORTH, C.K. and BAXTER, B.S. (1979), 'Fast FFT-based algorithm for phase estimation in speckle imaging', *Applied Optics*, Vol. 18, No. 12, 15 June, Pp. 2056–2061.

GARDEN, K.L. (1984), *An Overview of Computed Tomography*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

GARDENIER, P.H., LIM, C.A., TAN, D.G.H. and BATES, R.H.T. (1986), 'Aperture distribution phase from single radiation pattern measurement via Gerchberg-Saxton algorithm', *Electronics Letters*, Vol. 22, No. 2, January, Pp. 113–115.

GEHLBACH, S.M. and SOMMER, F.G. (1987), 'Frequency diversity speckle processing', *Ultrasonic Imaging*, Vol. 9, No. 2, April, Pp. 92–105.

GERCHBERG, R.W. (1974), 'Super-resolution through error energy reduction', *Optica Acta*, Vol. 21, No. 9, Pp. 709–720.

GERCHBERG, R.W. (1986), 'The lock problem in the Gerchberg-Saxton algorithm for phase retrieval', *Optik*, Vol. 74, No. 3, October, Pp. 91–93.

GERCHBERG, R.W. and SAXTON, W.O. (1972), 'A practical algorithm for the determination of phase from image and diffraction plane pictures', *Optik*, Vol. 35, No. 2, April, Pp. 237–246.

GEVINS, A.S. (1984), 'Analysis of the electromagnetic signals of the human brain: Milestones, obstacles, and goals.', *IEEE Transactions on Biomedical Engineering*, Vol. BME-31, No. 12, December, Pp. 833–850.

GEZARI, D.Y., LABEYRIE, A. and STACHNIK, R.V. (1972), 'Speckle interferometry: Diffraction-limited measurements of nine stars with the 200-inch telescope', *The Astrophysical Journal*, Vol. 173, No. 1, 1 April, Pp. L1–L5.

GLASER, I. (1987), 'Information processing with spatially incoherent light', In WOLF, E. (Ed.), *Progress in Optics*, North-Holland, Amsterdam, Chap. V, Pp. 389–509.

GLOVER, J.R., KTONAS, P.Y., RAGHAVAN, N., URUNELA, J.M., VELAMURI, S. and REILLY, E. (1986), 'A multichannel signal processor for the detection of epileptogenic sharp transients in the EEG', *IEEE Transactions on Biomedical Engineering*, Vol. BME-12, No. 12, December.

GLOVER, J., RAGHAVAN, N. and KTONAS, P. (1987), 'Context-based detection of epileptogenic EEG sharp transients', In *Proceedings of the IEEE/Ninth Annual Conference of the Engineering in Medicine and Biology Society*, IEEE, Boston, Pp. 1274–1275.

GONZALEZ, R.C. and WINTZ, P. (1977), *Digital Image Processing*, Addison-Wesley Publishing Company, Inc.

GOODMAN, J.W. (1968), *Fourier Optics*, McGraw-Hill, New York.

GOODMAN, J.W. (1976), 'Some fundamental properties of speckle', *Journal of the Optical Society of America*, Vol. 66, No. 11, November, Pp. 1145–1149.

GOODMAN, J.W. and BELSHER, J.F. (1976), 'Fundamental limitations in linear invariant restoration of atmospherically degraded images', In *Proceedings SPIE: Vol. 75: Imaging Through the Atmosphere*, 22-23 March, Pp. 243–254.

GOTMAN, J. (1985), 'Automatic recognition of interictal spikes', In GOTMAN, J., IVES, J.R. and GLOOR, P. (Eds.), *Long-term monitoring in epilepsy (EEG supplement number 37)*, Elsevier Science Publishers B.V. (Biomedical Division), Pp. 93–114.

GOTMAN, J. and GLOOR, P. (1976), 'Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG', *Electroencephalography and Clinical Neurophysiology*, Vol. 41, Pp. 513–529.

GRANRATH, D.J. (1987), 'Towards a fast triple correlation algorithm', In IDELL, P.S. (Ed.), *Proceedings SPIE: Vol. 828: Digital Image Recovery and Synthesis*, SPIE, 17–18 August, Pp. 95–100.

GULL, S.F. and DANIELL, G.J. (1978), 'Image reconstruction from incomplete and noisy data', *Nature*, Vol. 272, Pp. 686–690.

HAGER, W.W. (1987), 'A modified fast Fourier transform for polynomial evaluation and the Jenkins-Traub algorithm', *Numerische Matematik*, Vol. 50, Pp. 253–261.

HAWKES, P.W. (Ed.) (1980), *Computer Processing of Electron Microscope Images*, Vol. 13 of Topics in Current Physics, Springer-Verlag, Berlin.

HAYES, M.H. (1982), 'The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-30, Pp. 140–154.

HAYES, M.H. and M^{c}CLELLAN, J.H. (1982), 'Reducible polynomials in more than one variable', *Proceedings IEEE*, Vol. 70, No. 2, February, Pp. 197–198.

HAYES, M.H., LIM, J.S. and OPPENHEIM, A.V. (1980), 'Signal reconstruction from phase or magnitude', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-28, Pp. 672–680.

HAYES-ROTH, F. (1985), 'Rule-based systems', *Communications of the Association of Computing Machinery*, Vol. 28, Pp. 921–932.

HAYKIN, S. (1983), *Communication Systems*, John Wiley & Sons, New York, 2 ed.

HEDLEY ROBINSON, J. (1972), *Astronomy Data Book*, David and Charles (Holdings) Limited, Newton Abbot, Devon.

HEGE, E.K., HUBBARD, E.N., DRUMMOND, J.D., STRITTMATTER, P.A., WORDEN, S.P. and LAUER, T. (1982), 'Speckle interferometric observations of Pluto and Charon', *Icarus*, Vol. 50, No. 1, April, Pp. 72–81.

HERMAN, G.T. (Ed.) (1979), *Image Reconstruction from Projections: Implementations and Applications*, Vol. 32 of Topics in Applied Physics, Springer-Verlag, Berlin.

HILL, A.G. and TOWNSEND, H. (1973), 'The automatic estimation of epileptic spike activity', *International Journal of Biomedical Computing*, Vol. 4, Pp. 149–156.

HOFMANN, K. and WEIGELT, G. (1987), 'Astronomical speckle masking: Image reconstruction by cross triple correlation', *Applied Optics*, Vol. 26, No. 10, 15 May, Pp. 2011–2015.

HÖGBOM, J.A. (1974), 'Aperture synthesis with a non-regular distribution of interferometer baselines', *Astronomy and Astrophysics Supplementry Series*, Vol. 15, No. 3, June, Pp. 417–426.

HOWELL, R.R., M^{c}CARTHY, D.W. and LOW, F.J. (1981), 'One-dimensional infrared speckle interferometry', *The Astrophysical Journal Letters*, Vol. 251, 1 December, Pp. L21–L25.

HUANG, T.S. (Ed.) (1981), *Two-dimensional Digital Signal Processing I: Linear Filters*, Vol. 42 of Topics in applied physics, Springer-Verlag, Berlin.

HUNT, B.R. (1979), 'Matrix formulation of the reconstruction of phase values from phase differences', *Journal of the Optical Society of America*, Vol. 69, No. 3, March, Pp. 393–399.

HUNT, B.R., FRIGHT, W.R. and BATES, R.H.T. (1983), 'Analysis of the shift-and-add method for imaging through turbulent media', *Journal of the Optical Society of America*, Vol. 73, No. 4, April, Pp. 456–465.

IZRAELEVITZ, D. and LIM, J.S. (1987), 'A new direct algorithm for image reconstruction from Fourier transform magnitude', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-35, Pp. 511–519.

JAYNES, E.T. (1982), 'On the rationale of maximum-entropy methods', *Proceedings IEEE*, Vol. 70, No. 9, September, Pp. 939–952.

JEFFREYS, W.H. (1980), 'On the method of least squares', *The Astronomical Journal*, Vol. 85, No. 2, February, Pp. 177–181.

JENKINS, M.A. and TRAUB, J.F. (1972), 'Algorithm 419 — zeros of a complex polynomial', *Communications of the Association of Computing Machinery*, Vol. 15, No. 2, February, Pp. 97–99.

JENNISON, R.C. (1958), 'A phase sensitive interferometric technique for the measurement of Fourier transforms of spatial brightness distributions of small angular extent', *Monthly Notices of the Royal Astronomical Society*, Vol. 118, Pp. 276–284.

JOHNSON, R. and AITKEN, G.J.M. (1989), 'Statistical properties of the photon address, phase-gradient algorithm', *Journal of the Optical Society of America A*, Vol. 6, No. 1, January, Pp. 56–61.

KAK, A.C. (1979), 'Computerized tomography with X-ray, emission and ultrasound sources', *Proceedings IEEE*, Vol. 67, No. 9, September, Pp. 1245–1271.

KANI, L.M. and DAINTY, J.C. (1988), 'Super-resolution using the Gerchberg algorithm', *Optics Communications*, Vol. 68, No. 1, 1 September, Pp. 11–17.

KING, K.F. and MORAN, P.R. (1984), 'A unified description of NMR imaging, data-collection strategies, and reconstruction', *Medical Physics*, Vol. 11, No. 1, January/February, Pp. 1–14.

KIRKPATRICK, S., GELATT, C.D. and VECCHI, M.P. (1983), 'Optimization by simulated annealing', *Science*, Vol. 220, Pp. 671–680.

KNOX, K.T. (1976), 'Image retrieval from astronomical speckle patterns', *Journal of the Optical Society of America*, Vol. 66, No. 11, November, Pp. 1236–1239.

KNOX, K.T. and THOMPSON, B.J. (1974), 'Recovery of images from atmospherically degraded short-exposure photographs', *The Astrophysical Journal Letters*, Vol. 193, No. 1, 1 October, Pp. L45–L48.

KORFF, D., DRYDEN, G. and MILLER, M.G. (1972), 'Information retrieval from atmospheric induced speckle patterns', *Optics Communications*, Vol. 5, No. 3, June, Pp. 187–192.

KREYSIG, E. (1979), *Advanced Engineering Mathematics*, John Wiley and Sons, New York.

KTONAS, P.Y. (1983), 'Automated analysis of abnormal electroencephalograms', *CRC Critical Reviews in Biomedical Engineering*, Vol. 9, No. 1, Pp. 39–97.

KTONAS, P.Y., LUOH, W.M., KEJARIWAL, M.L., REILLY, E.L. and SEWARD, M.A. (1981), 'Computer-aided quantification of EEG spike and sharp wave characteristics', *Electroencephalography and Clinical Neurophysiology*, Vol. 51, Pp. 237–243.

LABEYRIE, A. (1970), 'Attainment of diffraction-limited resolution in large telescopes by Fourier analysing speckle patterns in star images', *Astronomy and Astrophysics*, Vol. 6, No. 1, May, Pp. 85–87.

LABEYRIE, A. (1976), 'High-resolution techniques in optical astronomy', In WOLF, E. (Ed.), *Progress in Optics, Vol. 9*, North-Holland, Amsterdam, Chap. 7, Pp. 47–87.

LANE, R.G. (1987), 'Recovery of complex images from Fourier magnitude', *Optics Communications*, Vol. 63, No. 1, 1 July, Pp. 6–10.

LANE, R.G. (1988), *Blind Deconvolution and Phase Retrieval*, PhD thesis, University of Canterbury, Christchurch, New Zealand, March.

LANE, R.G. and BATES, R.H.T. (1987a), 'Automatic multi-dimensional deconvolution', *Journal of the Optical Society of America A*, Vol. 4, No. 1, January, Pp. 180–188.

LANE, R.G. and BATES, R.H.T. (1987b), 'Relevance for blind deconvolution of recovering Fourier magnitude', *Optics Communications*, Vol. 63, No. 1, 1 July, Pp. 11–14.

LANE, R.G., FRIGHT, W.R. and BATES, R.H.T. (1987), 'Direct phase retrieval', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-35, Pp. 520–526.

LAWTON, W. and MORRISON, J. (1987), 'Factoring trigonometric polynomials regarded as entire functions of exponential type', *Journal of the Optical Society of America A*, Vol. 4, No. 1, January, Pp. 105–111.

LEINERT, C. and DYCK, H.M. (1983), 'Speckle interferometry degraded by irregular motion of a scanning telescope mirror', *Applied Optics*, Vol. 22, No. 16, Pp. 2403–2404.

LESEM, L.B., HIRSH, P.M. and JORDAN, J.A. (1969), 'The kinoform: A new wavefront reconstruction device', *IBM Journal of Research and Development*, Vol. 13, Pp. 150–155.

LEVI, A. and STARK, H. (1984), 'Image restoration by the method of generalised projections with application to restoration from magnitude', *Journal of the Optical Society of America A*, Vol. 1, Pp. 932–943.

LEVIN, B.J. (1964), *Distribution of Zeros of Entire Functions*, American Mathematical Society, Providence, Rhode Island.

LEWITT, R.M. and BATES, R.H.T. (1978), 'Image reconstruction from projections: I: General theoretical considerations', *Optik*, Vol. 50, No. 1, February, Pp. 19–33.

LEWITT, R.M., BATES, R.H.T. and PETERS, T.M. (1978), 'Image reconstruction from projections: II: Modified back-projection methods', *Optik*, Vol. 50, No. 2, March, Pp. 85–109.

LOHMANN, A.W. (1986), 'Shift and rotation tolerant image recovery from triple correlations', *Optik*, Vol. 73, No. 3, Pp. 127–131.

LOHMANN, A.W. (1988), 'Pattern recognition based on triple correlation', *Optik*, Vol. 78, No. 3, Pp. 117–120.

LOHMANN, A.W. and WIRNITZER, B. (1984), 'Triple correlations', *Proceedings IEEE*, Vol. 72, No. 7, July, Pp. 889–901.

LOHMANN, A.W., WEIGELT, G. and WIRNITZER, B. (1983), 'Speckle masking in astronomy: Triple correlation theory and applications', *Applied Optics*, Vol. 22, No. 24, 15 December, Pp. 4028–4037.

LYNDS, C.R., WORDEN, S.P. and HARVEY, J.W. (1976), 'Digital image reconstruction applied to Alpha Orionis', *The Astrophysical Journal*, Vol. 207, No. 1, 1 July, Pp. 174–180. Plates 3, 4.

MAKHOUL, J. (1975), 'Linear prediction: A tutorial review', *Proceedings IEEE*, Vol. 63, No. 4, April, Pp. 561–580.

MARIOTTI, J.M. (1983), 'Experimental results on atmospheric turbulence obtained with an infrared speckle-interferometer', *Optica Acta*, Vol. 30, No. 6, June, Pp. 831–840.

MARIOTTI, J.M., CHELLI, A., FOY, R., LÉNA, P., SIBILLE, F. and TCHOUNTONOV, G. (1983), 'Infrared speckle imaging: Improvement of the method; results on Miras and protostars', *Astronomy and Astrophysics*, Vol. 120, No. 2, April, Pp. 237–248.

MARKEL, J.D. and GRAY, JR., A.H. (1976), *Linear Prediction of Speech*, Vol. 12 of Communications and Cybernetics, Springer-Verlag, Berlin.

McALISTER, H.A. (1985), 'High angular resolution measurements of stellar properties', *Annual Review of Astronomy and Astrophysics*, Vol. 23, Pp. 59–87.

McALISTER, H.A. and HARTKOPF, W.I. (1988), 'Second catelog of interferometric measurements of binary stars', CHARA Contribution No. 2, Centre for high angular resolution astronomy, Georgia State University, Atlanta, USA., October.

McCALLUM, B.C. (1989), PhD Thesis. In preparation, University of Canterbury, Christchurch, New Zealand.

McCALLUM, B.C. and BATES, R.H.T. (1989), 'Towards a strategy for automatic phase retrieval from noisy Fourier intensities', *Journal of Modern Optics*, Vol. 36.

McKINNON, A.E., McDONNELL, M.J., NAPIER, P.J. and BATES, R.H.T. (1976), 'Self-consistent deconvolution: II: Applications', *Optik*, Vol. 44, No. 3, February, Pp. 253–272.

McLEAN, I.S. (1988), 'Infrared astronomy's new image', *Sky and Telescope*, Vol. 75, No. 3, Pp. 254–258.

METROPOLIS, N., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A. and TELLER, E. (1953), 'Equation of state calculations by fast computing machines', *Journal of Chemical Physics*, Vol. 21, Pp. 1087–1092.

MILLER, R.L. (1959), 'Nature of the vocal cord wave', *Journal of the Acoustical Society of America*, Vol. 31, Pp. 667–677.

MILNER, M.O. (1979), *Error Correction in Images and Imaging Instruments*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

MINARD, R.A. (1985), *Imaging in a Distorting Medium*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

MINARD, R.A., ROBINSON, B.S. and BATES, R.H.T. (1985), 'Full-wave computed tomography part 3: Coherent shift-and-add imaging', *Proceedings IEE, Part A*, Vol. 132, No. 1, January, Pp. 50–58.

MISSELL, D.L. (1978), 'The phase problem in electron microscopy', In COSSLETT, V.E. and BARER, R. (Eds.), *Advances in Optical and Electron Microscopy*, Academic, Pp. 185–279.

MONIN, J.L., VAUGLIN, I., SIBILLE, F. and AUDAIRE, L. (1987), 'A new infrared camera for the 2–5μm range', *Astronomy and Astrophysics*, Vol. 172, Pp. 368–374.

MORRIS, D. (1985), 'Phase retrieval in the radio holography of reflector antennas and radio telescopes', *IEEE Transactions on Antennas and Propagation*, Vol. AP-33, Pp. 749–755.

McCARTHY, D.W. and COBB, M.L. (1986), 'Comparative study of deconvolution techniques for infrared speckle interferometry', In *Proceedings SPIE: Vol.* 627, SPIE, P. 797.

McCARTHY, D.W., COBB, M.L. and PROBST, R.G. (1987), 'Gliese 866: A new, low-mass binary in the solar neighborhood', *The Astronomical Journal*, Vol. 93, No. 6, June, Pp. 1535–1538.

MULLER, R.A. and BUFFINGTON, A. (1974), 'Real-time correction of atmospherically degraded telescope images through image sharpening', *Journal of the Optical Society of America*, Vol. 64, No. 9, September, Pp. 1200–1210.

NAKAJIMA, N. (1988), 'Phase retrieval using the logarithmic Hilbert transform and the Fourier-series expansion', *Journal of the Optical Society of America A*, Vol. 5, No. 2, February, Pp. 257–262.

NAKAJIMA, N. and ASAKURA, T. (1983), 'Extraction of the influence of zeros from the image intensity in the phase retrieval using the logarithmic Hilbert transform', *Optik*, Vol. 63, No. 2, January, Pp. 99–108.

NAPIER, P.J. (1971), *Reconstruction of Radiating Sources*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

NAPIER, P.J. and BATES, R.H.T. (1971), 'Holographic approach to radiation pattern measurement — II. Experimental verification', *International Journal of Engineering Science*, Vol. 9, No. 12, December, Pp. 1193–1208.

NAPIER, P.J. and BATES, R.H.T. (1974), 'Inferring phase information from modulus information in two-dimensional aperture synthesis', *Astronomy and Astrophysics Supplementry Series*, Vol. 15, No. 3, June, Pp. 427–430.

NAPIER, P.J., THOMPSON, A.R. and EKERS, R.D. (1983), 'The very large array: Design and performance of a modern synthesis radio telescope', *Proceedings IEEE*, Vol. 71, P. 1295.

NARAYAN, N. (1987), 'Phase retrieval with the maximum entropy method', In *Proceedings of the Joint Workshop on High-Resolution Imaging from the Ground Using Interferometric Techniques*, Oracle, Arizona, January 12-15, Pp. 183–186.

NIETO-VESPERINAS, M. and DAINTY, J.C. (1986), 'Phase recovery for two-dimensional digital objects by polynomial factorisation', *Optics Communications*, Vol. 58, No. 2, 15 May, Pp. 83–88.

NIETO-VESPERINAS, M. and MENDEZ, J.A. (1986), 'Phase retrieval by Monte-Carlo methods', *Optics Communications*, Vol. 59, No. 4, 15 September, Pp. 249–254.

NIETO-VESPERINAS, M., NAVARRO, R. and FUENTES, F.J. (1988), 'Performance of a simulated-annealing algorithm for phase retrieval', *Journal of the Optical Society of America A*, Vol. 5, No. 1, January, Pp. 30–38.

NITYANANDA, R. and NARAYAN, R. (1982), 'Maximum entropy image reconstruction - a practical non-information-theoretic approach', *Journal Astrophysics and Astronomy*, Vol. 3, Pp. 419–450.

NORTHCOTT, M.J., AYERS, G.R. and DAINTY, J.C. (1988), 'Algorithms for image reconstruction from photon-limited data using the triple correlation', *Journal of the Optical Society of America A*, Vol. 5, No. 7, July, Pp. 986–992.

OPPENHEIM, A.V. and LIM, J.S. (1981), 'The importance of phase in signals', *Proceedings IEEE*, Vol. 69, Pp. 529–541.

OPPENHEIM, A.V. and SCHAFER, R.W. (1975), *Digital Signal Processing*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

OPPENHEIM, A.V., SCHAFER, R.W. and STOCKHAM, JR., T.G. (1968), 'Non-linear filtering of multiplied and convolved signals', *Proceedings IEEE*, Vol. 56, No. 8, August, Pp. 1264–1291.

OPPENHEIM, A.V., HAYES, M.H. and LIM, J.S. (1982), 'Iterative procedures for signal reconstruction from Fourier transform phase', *Optical Engineering*, Vol. 21, Pp. 122–127.

PERI, D. (1987), 'Optical implementation of a phase retrieval algorithm', *Applied Optics*, Vol. 26, No. 9, 1 May, Pp. 1782–1785.

PROBST, R.G. (1977), 'Parallax, orbit, and mass of Ross 614', *The Astronomical Journal*, Vol. 82, No. 8, August, Pp. 656–661.

PROC. IEEE/EMBS (1987), *Proceedings of the IEEE/Ninth Annual Conference of the Engineering in Medicine and Biology Society*, IEEE, Boston.

RAMACHANDRAN, G.N. and SRINIVASAN, R. (1970), *Fourier Methods in Crystallography*, Wiley-Interscience, New York.

RAMIREZ, R.W. (1985), *The FFT: Fundamentals and Concepts*, Prentice-Hall, Englewood Cliffs, New Jersey.

READHEAD, A.C.S., WALKER, R.C., PEARSON, T.J. and COHEN, M.H. (1980), 'Mapping radio sources with uncalibrated visibility data', *Nature*, Vol. 285, Pp. 137–140.

REQUICHA, A.A.G. (1980), 'The zeros of entire functions: Theory and engineering applications', *Proceedings IEEE*, Vol. 68, No. 3, March, Pp. 308–328.

RIBAK, E. (1986), 'Astronomical imaging by filtered weighted-shift-and-add technique', *Journal of the Optical Society of America A*, Vol. 3, No. 12, December, Pp. 2069–2076.

RIBAK, E., HEGE, E.K. and CHRISTOU, J.C. (1985), 'Use of matched filtering to identify speckle locations', In ARSENAULT, H.H. (Ed.), *Proceedings SPIE: Vol. 556: International Conference on Speckle*, SPIE, August 20-23, Pp. 196–201.

ROBINSON, B.S. (1982), *Speckle Processing for Ultrasonic Imaging*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

RODDIER, F. (1981), 'The effects of atmospheric turbulence in optical astronomy', In WOLF, E. (Ed.), *Progress in Optics*, North-Holland, Pp. 283–376.

RODDIER, F. (1988), 'Interferometric imaging in optical astronomy', *Physics Reports*, Vol. 170, No. 2, November, Pp. 97–166.

ROSENFELD, A. and KAK, A.C. (1982), *Digital Picture Processing*, Vol. 1 of Computer Science and Applied Mathematics, Academic Press, New York, 2 ed.

ROY, A.E. and CLARKE, D. (1977), *Astronomy: Structure of the Universe*, Adam Hilger Ltd, Bristol, Bristol.

SANZ, J.L.C. and HUANG, T.S. (1984), 'Support-limited signal and image extrapolation', In HUANG, T.S. (Ed.), *Advances in Computer Vision and Image Processing: Vol. 1 — Image Reconstruction from Incomplete Observations*, JAI Press Inc., Greenwich, Connecticut, Chap. 1, Pp. 1–82.

SANZ, J.L.C. and HUANG, T.S. (1985), 'Polynomial system of equations and its applications to the study of the effect of noise on multidimensional Fourier transform phase retrieval from magnitude', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-33, Pp. 997–1004.

SAXTON, W.O. (1978), *Computer Techniques for Image Processing in Electron Microscopy*, Vol. 10 of Advances in electronics and electron physics, Academic Press, New York.

SCHAFER, R.W., MERSEREAU, R.M. and RICHARDS, M.A. (1981), 'Constrained iterative deconvolution algorithms', *Proceedings IEEE*, Vol. 69, Pp. 432–450.

SIBILLE, F., CHELLI, A. and LÉNA, P. (1979), 'Infrared speckle interferometry', *Astronomy and Astrophysics*, Vol. 79, No. 3, November, Pp. 315–328.

SINTON, A.M. (1986), *Contributions to Astronomical and Medical Information Processing*, PhD thesis, University of Canterbury, Christchurch, New Zealand, May.

SINTON, A.M., DAVEY, B.L.K. and BATES, R.H.T. (1986), 'Augmenting shift-and-add with zero-and-add', *Journal of the Optical Society of America A*, Vol. 3, No. 7, July, Pp. 1010–1017.

SPEHLMANN, R. (1981), *EEG Primer*, Elsevier/North Holland Biomedical Press, Amsterdam.

STANLEY, W.D. (1975), *Digital Signal Processing*, Reston Publishing Company, Inc., Reston, Virginia.

STEFANESCU, I.S. (1985), 'On the phase retrieval problem in two dimensions', *Journal Mathematical Physics*, Vol. 26, Pp. 2141–2160.

STOCKHAM, JR., T.G. (1971), 'Restoration of old acoustic recordings by means of digital signal processing', Audio Engineering Society Convention Abstracts.

STOCKHAM, JR., T.G., CANNON, T.M. and INGEBRETSON, R.B. (1975), 'Blind deconvolution through digital signal processing', *Proceedings IEEE*, Vol. 63, No. 4, April, Pp. 678–692.

STROHBEHN, J.W. (1971), 'Optical propagation through the turbulent atmosphere', In WOLF, E. (Ed.), *Progress in Optics*, North-Holland, Amsterdam, Pp. 75–122.

THOMPSON, A.R., MORAN, J.M. and SWENSON, G.W. (1986), *Interferometry and Synthesis in Radio Astronomy*, John Wiley and Sons, New York.

TITCHMARSH, E.C. (1932), *The Theory of Functions*, Clarendon Press, Oxford.

TOM, V.T., QUATIERI, T.F., HAYES, M.H. and MCCLELLAN, J.H. (1981), 'Convergence of non-expansive signal reconstruction algorithms', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-29.

TSAO, J. and STEINBERG, B.D. (1988), 'Reduction of sidelobe and speckle artifacts in microwave imaging: the CLEAN technique', *IEEE Transactions on Antennas and Propagation*, Vol. 36, No. 4, April, P. 543.

ULRYCH, T. (1971), 'Application of homomorphic deconvolution to seismology', *Geophysics*, Vol. 36, No. 4, August, Pp. 650–660.

VON DER LÜHE, O. (1984), 'Estimating Fried's parameter from a time series of an arbitrary resolved object imaged through atmospheric turbulence', *Journal of the Optical Society of America A*, Vol. 1, No. 5, May, Pp. 510–519.

VON DER LÜHE, O. (1988), 'Signal transfer function of the Knox-Thompson speckle imaging technique', *Journal of the Optical Society of America A*, Vol. 5, No. 5, May, Pp. 721–729.

WALKER, R.J. (1950), *Algebraic Curves*, Princeton University Press, Princeton, New Jersey.

WALKER, J.G. (1981), 'Object reconstruction from turbulence-degraded images', *Optica Acta*, Vol. 28, No. 8, August, Pp. 1017–1019.

WALTHER, A. (1963), 'The question of phase retrieval in optics', *Optica Acta*, Vol. 11, Pp. 41–49.

WANG, D.L. and LIM, J.S. (1982), 'The unimportance of phase in speech enhancement', *IEEE Transactions on Acoustics Speech and Signal Processing*, Vol. ASSP-30, Pp. 679–681.

WEIGELT, G. and BAIER, G. (1985), 'R136a in the 30 Doradus nebula resolved by holographic speckle interferometry', *Astronomy and Astrophysics Letters*, Vol. 150, No. 1, September, Pp. L18–L20.

WEIGELT, G. and WIRNITZER, B. (1983), 'Image reconstruction by the speckle-masking method', *Optics Letters*, Vol. 8, No. 7, July, Pp. 389–391.

WELLS, P.N.T. (Ed.) (1982), *Scientific Basis of Medical Imaging*, Churchill Livingstone, Medical Divison of Longman Group Limited, Robert Stevenson House, Edinburgh.

WELTER, G.L. and WORDEN, S.P. (1978), 'A method for processing stellar speckle interferometry data', *Journal of the Optical Society of America*, Vol. 68, No. 9, September, Pp. 1271–1275.

WIRNITZER, B. (1985), 'Bispectral analysis at low light levels and astronomical speckle masking', *Journal of the Optical Society of America A*, Vol. 2, No. 1, January, Pp. 14–21.

WOOLF, N.J. (1982), 'High resolution imaging from the ground', In BURBRIDGE, G. (Ed.), *Annual Review of Astronomy and Astrophysics*, Annual Reviews Inc., Palo Alto, California, Pp. 367–398.

YOULA, D.C. and WEBB, H. (1982), 'Image restoration by the method of convex projections: Part 1 - Theory', *IEEE Transactions on Medical Imaging*, Vol. MI-1, Pp. 81–94.